Evaluation of Candidate Rain Gages For Upgrading Precipitation Measurement Tools For the National Atmospheric Deposition Program

Water-Resources Investigations Report 02–4302

U.S. Department of the Interior
U.S. Geological Survey

Much appreciation is extended to the following U.S. Geological Survey employees for their contributions to this report:
Paul Tippet, Hydrologic Instrumentation Facility, for overseeing Phase I and Phase II gage testing; members of the Colorado District Reports Production Unit (Alene Brogan, Robert Olmstead, and John Evans) for final production of the report; and Mary Kidd for her editorial and verification reviews.

Appreciation also is extended to the technical reviewers:
Dr. James Lynch, Penn State University; Jeff Cole, National Center for Atmospheric Research; and
Roger Claybrooke, Illinois State Water Survey.

Evaluation of Candidate Rain Gages for Upgrading Precipitation Measurement Tools for the National Atmospheric Deposition Program

By John D. Gordon

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 02–4302

Denver, Colorado
2003

U.S. DEPARTMENT OF THE INTERIOR
GALE A. NORTON, Secretary

U.S. GEOLOGICAL SURVEY
Charles G. Groat, Director

For additional information write to:

Chief, Branch of Quality Systems
U.S. Geological Survey
Box 25046, Mail Stop 401
Denver Federal Center
Denver, CO 80225-0046

Copies of this report can be purchased from:

U.S. Geological Survey
Information Services
Box 25286
Denver Federal Center
Denver, CO 80225

CONTENTS

FIGURES

TABLES

CONVERSION FACTORS

Multiply	By	To obtain
inch	2.540	centimeter
quart	0.946	liter
ounce, fluid	29.577	milliliter
inch	25.400	millimeter
meter	3.281	foot
kilogram	2.2045	pound

Temperature in degrees Celsius (° C) may be converted to degrees Fahrenheit (° F) or degrees Celsius to degrees Fahrenheit by using the following:

$$°F = 9/5\ (°C) + 32$$

$$°C = 5/9\ (°F - 32)$$

ADDITIONAL ABBREVIATION AND SYMBOL

probability (p)
maximum probability of rejecting the null hypothesis when it is true (α)

Evaluation of Candidate Rain Gages for Upgrading Precipitation Measurement Tools for the National Atmospheric Deposition Program

By John D. Gordon

Abstract

The National Atmospheric Deposition Program (NADP) was established in 1977 to investigate atmospheric deposition and its effects on the environment. Since its establishment, precipitation records have been obtained at all NADP sites using a gage developed approximately 50 years ago—the Belfort 5–780 mechanical rain gage. In 1998 and 1999, a study was done by the U.S. Geological Survey to evaluate four recently developed, technologically advanced rain gages as possible replacement candidates for the mechanical gage currently (2002) in use by the NADP. The gage types evaluated were the Belfort 3200, Geonor T–200, ETI Noah II, and the OTT PLUVIO. The Belfort 5–780 was included in the study to compare the performance of the rain gage currently (2002) used by NADP to the performance of the more recently developed gages. As a reference gage, the NovaLynx Model 260–2510 National Weather Service type stick gage also was included in the study. Two individual gages of each type were included in the study to evaluate precision between gages of the same type. A two-phase evaluation was completed. Phase I consisted of indoor bench tests with known amounts of simulated rainfall applied in 20 individual tests. Phase II consisted of outdoor testing by collecting precipitation during a 26-week period near Bay St. Louis, Mississippi. The ETI Noah II, OTT PLUVIO, and NovaLynx stick gages consistently recorded depths more commensurate with the amounts of applied simulated rainfall in Phase I testing than the Geonor T–200, Belfort 5–780, and Belfort 3200 gages. Gages where both the median difference between the measured and applied simulated rainfall and the interquartile range of all of their measured minus applied simulated rainfall differences were small (less than or equal to 0.01 inch) were judged to have performed very well in Phase I testing. The median and interquartile-range values were 0.01 inch or less for each of the ETI Noah II gages, OTT PLUVIO gages, and NovaLynx stick gages. The performance of the Geonor T–200 and Belfort 3200 gages was affected by technical problems during Phase I testing. As part of the evaluation of Phase II results, the average weekly precipitation totals obtained from the Belfort 5–780 gages and from each of the gages under consideration as possible replacements for the Belfort 5–780 gage were all compared with the average precipitation weekly totals obtained from two NovaLynx stick gages. The median absolute differences between a particular gage model and the NovaLynx stick reference gage for the 26 weeks of outdoor testing ranged from 0.04 inch for the ETI Noah II and OTT PLUVIO gages to 0.06 inch for the Geonor T–200. The total absolute difference between a particular gage type and the reference gage ranged from 1.23 inches for the Belfort 5–780 to 1.83 inches for the Geonor T–200 gages. Because the Belfort 3200 gages were inoperable for most of the Phase II testing, it is not meaningful to include the results from that gage type in a calculation of median or total absolute differences. The OTT PLUVIO proved to be the most reliable gage in Phase I and II testing, operating trouble free over the duration of the study.

INTRODUCTION

Since 1977, the National Atmospheric Deposition Program (NADP) has investigated atmospheric deposition and its effects on the environment. Currently (2002), the NADP operates approximately 260 monitoring sites at approximately 230 locations throughout the United States, and in Puerto Rico and the Virgin Islands (fig. 1), and at a location in Canada. Three separate networks comprise the NADP—the National Trends Network (NTN), the Mercury Deposition Network (MDN), and the Atmospheric Integrated Research and Monitoring Network (AIRMoN). The NTN is a weekly monitoring network designed to collect data on the chemistry of precipitation for monitoring geographical and temporal long-term trends. The objective of the MDN is to develop a national database of weekly concentrations of total mercury in precipitation and the seasonal and annual flux of total mercury in wet deposition. AIRMoN is a daily precipitation-chemistry-monitoring network sponsored by the National Oceanic and Atmospheric Administration Air Resources Laboratory (National Oceanic and Atmospheric Administration, 2002). Whereas the NTN was designed to characterize long-term trends in the chemical climate of the United States, AIRMoN was designed to provide data with a greater temporal resolution.

All three NADP networks use the same rain gage, the Belfort 5–780, to determine precipitation depth. The Belfort 5–780 is a mechanical gage developed in the 1940's. In addition to being an older design, the gages used by the NADP are aging; most Belfort 5–780 gages in operation today have been continuously deployed in the field for more than 15 years, and in 1997 alone there were more than 180 equipment failures (Claybrooke and others, 2000). Because many of the rain gages at the NADP sites have been operated for nearly 2 decades, there are concerns about the continued reliability of these aging instruments.

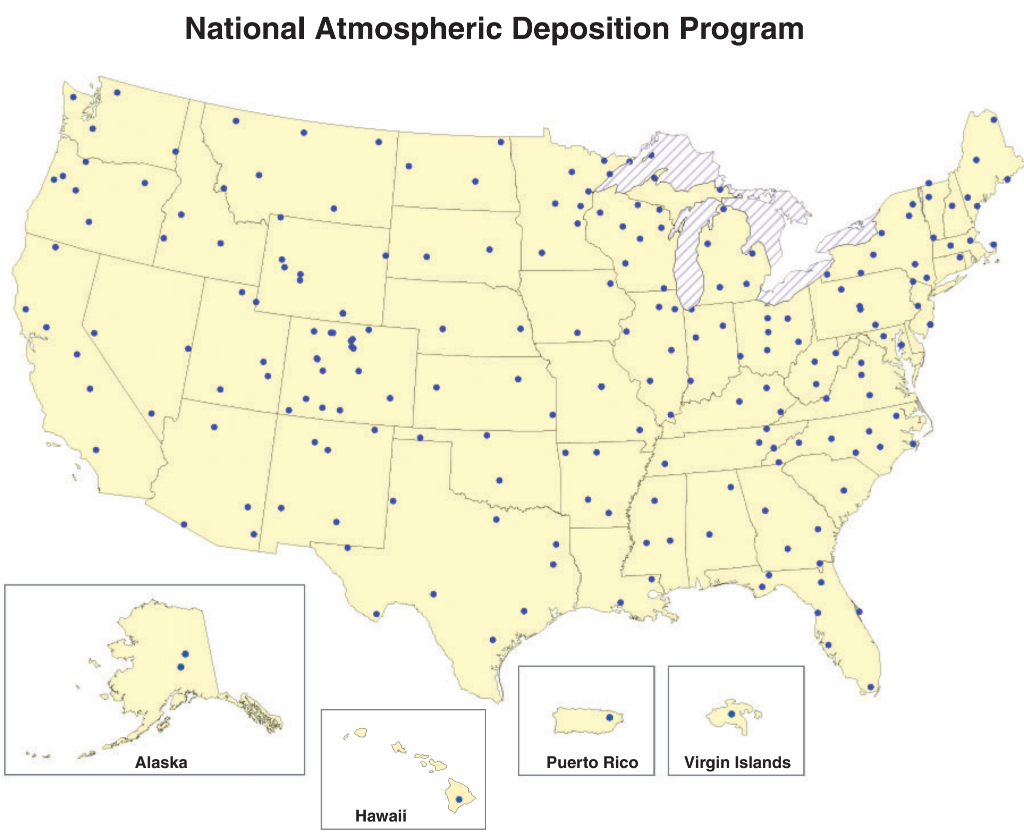


Figure 1. Location of National Atmospheric Deposition Program sites, December 2001.

The scientific community associated with the NADP, aware that there have been considerable advances in rain-gage technology, is actively exploring options for replacing the Belfort 5–780. Since the time the Belfort 5–780 gages were installed, many new precipitation-measurement tools have become available. In 1998 and 1999, the U.S. Geological Survey (USGS) evaluated the performance of several recently developed, technologically advanced rain gages as possible replacements for the Belfort 5–780. The ideal replacement gage would provide the NADP with precise and accurate data retrievable in a variety of electronic formats, with minimal human intervention and maintenance. This report details the evaluation of four possible replacement gages for upgrading precipitation measurement in the NADP. The accuracy, precision, and reliability of the replacement gage candidates also were evaluated.

PREVIOUS RAIN-GAGE STUDIES

Difficulties involved in obtaining accurate measurements of solid (snow/ice) and liquid (rain) precipitation have been recognized for many years. Numerous studies have been conducted to evaluate the accuracy and performance of rain gages. The results of these studies show that the type of collection method used can significantly affect the quantity of precipitation measured. Although the types of collection methods and the magnitude of measurement error differed in each study, many studies concluded that overcatch due to raindrop splash, undercatch due to wind velocity and shielding, and errors associated with the diameter of gage orifice (opening), amount of time the orifice was open, and the type of rain gage used contributed the most to measurement precision and accuracy (Emerson and Macek-Rowland, 1990; Goodison and others, 1981; Sturges, 1984; Yang and others, 1999).

Intensive research into rain-gage overcatch due to raindrop splash was conducted more than 70 years ago (Gold, 1931). Gold's research suggested that 2-mm raindrops impinging on a wet surface could splash to heights of 4.2 ft. A few years after Gold published his research, Ashmore (1934) reported that the highest splash would approach 4 ft on hard surfaces such as marble, asphalt, tarmac, concrete, and red bricks. Golubev (1985) found that the maximum flight distance of raindrop splash was 3.6 ft after impacting on wet or dry ceramics or on standing water. Golubev (1985) also found that only 2.7 percent of raindrop splash would exceed a height of 11.8 inches.

A major concern with the use of rain gages is the effect of wind on their accuracy. For example, gages with sloped surfaces tapering upward to the orifice can deflect and accelerate wind over the gage, reducing catch (Jeff Cole, National Center for Atmospheric Research, written commun., 2001). Wind-induced eddies around a gage will deflect precipitation that normally would be caught or may even eject drops that have already entered the collector (Handock, 1960). The higher the gage is installed above the ground surface, the greater the wind-induced errors (Handock, 1960). Turbulent eddies created by the wind in the gage mouth may reduce the catch, and the increased horizontal windspeed over the orifice may result in the transport of small droplets across the opening, which would otherwise, in undisturbed air flow, fall into the gage (Bruce and Potter, 1957). During the 1960's, scientists working at the Precipitation Polygon in Valdai, Russia, installed standard 8-inch United States nonrecording rain gages in 64.58-ft^2 clearings for a 5-year study of wetting, undercatch, and evaporation losses. Their findings indicate a 4-percent undercatch of liquid (rain) precipitation if gages were installed without shielding from wind effects (Golubev and others, 1992). Liquid precipitation undercatch from wind effects also has been estimated at about 4 percent by other researchers (Jeff Cole, National Center for Atmospheric Research, written commun., 2001). Gage size, shape, configuration, and location result in unique wind-deflection characteristics that can contribute to measurement errors. These gage attributes have a pronounced effect on wind turbulence over and around the gage and can substantially reduce gage catch efficiency. The diameter and construction of the gage orifice and the resulting effects on wind turbulence are less important than the size, shape, and location of the gage itself.

Key physical characteristics of the gage (diameter, construction, material, configuration, and size) also can contribute to rain-gage error. The diameter and construction of the gage orifice contribute to rain-gage error by retarding or enhancing wetting loss. Wetting loss is defined as water subject to evaporation from the surface of the inner walls of the precipitation gage after a precipitation event and water retained on the walls of the gage and its containers after its contents are emptied (Metcalfe and others, 1994).

While evaluating the performance of various technologically advanced rain gages, it is important to keep in mind that wind-induced gage undercatch of precipitation, among other known systematic errors, is the greatest source of bias in precipitation observation (Yang and others, 1999). Yang and others also referenced the work by Karl and others (1993), Groisman and Legates (1994), and Metcalfe and others (1997), commenting that changes in instrumentation may introduce a discontinuity into precipitation time series because the gage measurement is affected by gage design (that is, physical characteristics of the gage), including particularly whether the gage is equipped with a wind shield. According to Yang and others (1999), numerous studies (Larkin, 1947; Larson and Peck, 1974; Goodison and others, 1981; Sturges, 1984; and Hanson, 1989) have shown that a shielded gage (in locations with significant amounts of solid precipitation) can catch up to 50 percent more precipitation than its unshielded counterpart for the same environmental conditions. It is important to note that these studies were all done in areas where a significant component of the precipitation was snow. Studies where the precipitation is primarily rain consistently show much smaller errors associated with shielding. As the NADP considers updating its aging precipitation gages in order to obtain more precise data, it is even more advisable that rain-gage shielding be adopted as standard practice by the network. Currently, rain-gage shielding is optional at NADP sites. More NADP sites lack shielding than those that have shielding for their rain gages. Yang and others (1999) report that the combination of precipitation records from shielded and unshielded gages can result in inhomogeneous precipitation time series and can lead to incorrect spatial interpretations. Because time-series analyses on atmospheric deposition data collected by the NADP rely on shielded and unshielded rain gages, it can be inferred that the same type of errors may be occurring. Spatial interpolations are also adversely affected by the current optional shield policy of the NADP (Jim Lynch, Penn State University, written commun., 2002).

In 1998, the Illinois State Water Survey began testing the ETI Noah II precipitation gage as a potential replacement for the Belfort 5–780 gage. The ETI Noah II has virtually no moving parts and contains electronic filters that reduce false readings caused by wind, temperature, and evaporation. To increase the sensitivity of the Belfort 5–780 gage for measuring small precipitation amounts, the standard 8-inch-diameter orifice was replaced with a 12-inch orifice. Liquid depth measurements from both gages were compared with reference measurements from a NovaLynx Model 260–2510 National Weather Service type stick gage (NovaLynx stick gage) located at the site. The NovaLynx stick gages have no mechanical parts and are commonly used as reference gages with which the performance of other gages is compared. This experiment continued for 1 year, from July 1998 to July 1999. In that time period, 96 precipitation events were measured, including light and heavy rain and snow. According to a paired t-test and a Wilcoxon signed-rank test, the data obtained from the ETI Noah II and Belfort 5–780 measurements were not significantly different. Both gages captured significantly less precipitation ($p<0.01$) than the NovaLynx stick gage. Differences between the Belfort 5–780 or ETI Noah II gages and NovaLynx stick gages averaged about 0.01 inch per event at this well-maintained test site operated by experienced personnel of the Illinois State Water Survey. One problematic observation of the ETI Noah II gage is that it consistently recorded false positive responses that appear to be temperature related (Claybrooke and others, 2000).

GAGES EVALUATED IN THIS STUDY

Four candidate rain-gage types (Belfort 3200, ETI Noah II, Geonor T–200, and OTT PLUVIO) were evaluated as possible replacements for the Belfort 5–780 rain gages used in the NADP. The performance of the candidate rain gages were compared to that of the Belfort 5–780 and the Nova Lynx stick gage. The Nova Lynx stick gage is used as a reference gage in this study because it has no mechanical parts.

The USGS Hydrologic Instrumentation Facility (HIF) in Bay St. Louis, Miss. provided testing for six rain-gage types. A summary of the technical specifications for each gage is shown in table 1. Two units of each of the following models were selected for testing:

Belfort Hi-Capacity Precipitation Gage model 3200 (Belfort 3200)

Belfort Universal Precipitation Gage series 5–780 (Belfort 5–780)

Geonor Precipitation Gage T–200 (Geonor T–200)

ETI Noah II

OTT PLUVIO

NovaLynx Model 260–2510 National Weather Service type stick gage (NovaLynx stick gage)

Table 1. Specifications for each gage type used in the study

[n/a, not applicable]

Name	Orifice diameter, in inches	Orifice diameter, in millimeters	1-inch liquid equivalent	Gage body diameter, in inches	Gage body diameter, in millimeters	Gage height, in inches	Gage height, in millimeters	Gage weight, in pounds	Gage weight, in kilograms	Power required, in volts	Sensor type
Belfort 3200	12.01	305.05	1,853	20	508.00	48	1,219.2	83	37.65	12	Vibrating force transducer
Belfort 5–780	8.00	203.20	824	14	355.60	35.5	901.7	25	11.34	n/a	Mechanical
Geonor T–200	6.28	159.51	507	15.35	390.00	29.92	760	28.66	13	12	Vibrating-wire gage sensor
ETI Noah II	11.90	302.26	1,820	12	304.80	34	863.6	30	13.61	12	Load transducer
OTT PLUVIO	8.27	210.06	507	8.27	210.00	22.44	570	13.2	6.0	9 to 16	Vibrating-wire gage sensor
NovaLynx Model NovaLynx 260–2510 NWS type	8.00	203.20	841	8	203.20	27	685.8	7	3.2	n/a	n/a

[1]At 20 degrees Celsius.
[2]At 4 degrees Celsius.

Description of Gages

Belfort Hi-Capacity Precipitation Gage Model 3200

The Belfort 3200 (fig. 2) has a 12-inch-diameter orifice and a straight-sided collector to minimize wetting loss, snow bridging, and the effects of wind. Snow bridging is an effect that occurs primarily in wet snow conditions when snow clings to the sides of a precipitation gage and gradually accumulates until the gage orifice is capped with accumulated snow (Belfort Instrument Company, 2002b). The weight of collected precipitation is sensed through a vibrating force transducer. The system has a microprocessor built into the unit to aid field personnel with sampling, data filtering, and temperature compensation. A collector funnel is provided for liquid precipitation, and a drain cock is provided for fluid removal. Data are available in serial digital (RS232), analog voltage, current-loop, and simulated tipping-bucket formats. Belfort specifications state that the accuracy of the gage is 0.02 inch and the resolution or sensitivity is 0.01 inch. The unit can be powered with either 12-volt direct current or 115 or 230-volt alternating current (Belfort Instrument Company, 2002c).

Figure 2. Belfort 3200 rain gage.

Belfort Universal Precipitation Gage Series 5–780

The Belfort 5–780 is a mechanical gage with an 8-inch-diameter orifice. Precipitation enters the Belfort 5–780 (fig. 3) and travels through an aluminum funnel into a galvanized steel bucket. The collection bucket rests on a platform supported by a calibrated weighing spring. As the bucket fills with precipitation, the spring is depressed, causing the linked dual-traverse pens to be displaced. An inked trace is recorded on a paper strip chart wrapped around a rotating, clock-operated drum. A trained operator records the volume and timing of accumulated precipitation events on a 7-day strip chart that requires manual interpretation. Belfort recommends adding a known volume of oil to the bucket to retard evaporation. In the winter a known amount of antifreeze should be added to keep the water from freezing; the funnel also should be removed so that solid precipitation falls directly into the collecting bucket. Gage capacity of 12 inches (300 mm) is standard. Belfort specifies that the accuracy and sensitivity at full scale are 0.5 and 0.1 percent, respectively. Potentiometric output is available for transmitting data, which would eliminate the need for manual interpretation of strip charts. In its standard configuration, the Belfort 5–780 gage does not require power from an external source (Belfort Instrument Company, 2002a). The two Belfort 5–780 gages used in this study were more than 10 years old and had been used by the NADP for many years. To represent a comparison of up-to-date gage technology to the current gages used by the NADP, the Belfort 5–780 gages used in this study were of similar condition to the Belfort 5–780's currently deployed by the NADP.

Figure 3. Belfort 5–780 rain gage.

Geonor Precipitation Gage T–200

The Geonor Precipitation Gage T–200 (fig. 4) uses an electrically excited, vibrating-wire gage sensor (700–3,500 hertz) to generate a frequency output as a function of applied tension or weight. There are no mechanical moving parts in this gage; therefore, the gage should normally require very little onsite maintenance. The accuracy of the gage is 0.1 percent of full scale whereas the sensitivity is 0.004 inch. Data can be recorded at any desired interval, and the gage interfaces with most data-acquisition systems. For example, during this study, personnel at the HIF programmed the T–200 to take measurements every 15 minutes. Power to the unit is supplied by 12-volt direct current through the datalogger connected to the gage. To facilitate data transmission, special hybrid-circuitry cards are supplied to most retrieval systems (Geonor, Inc., 2002).

Figure 4. Geonor Precipitation Gage T–200.

ETI Noah II

The ETI Noah II (fig. 5) uses a Weight Measurement Assembly (WMA) sensor to detect the amount of collected precipitation. The WMA sensor is a stabilized amplifier that converts direct current impulses to alternating current, which then provides excitation to a load transducer. Output from the load transducer is fed to a 16-bit-resolution analog-to-digital converter, which provides an internal sensitivity of greater than 0.005 inch. The air temperature and sample weight are measured every 10 seconds. Additional software converts the 10-second values to 1-minute values, which are then summed into 15-minute values. The ETI Noah II provides onsite communication with remotely located sensors through cellular and land-line telephone systems, Geostationary Operational Environmental Satellite (GOES) communication, meteor-burst communications, and LEO (low-earth orbit) satellite data-retrieval systems. The ETI Noah II gage operates on 12-volt direct current (ETI Instrument Systems, Inc., 2002).

Figure 5. ETI Noah II rain gage.

OTT PLUVIO

The OTT PLUVIO gage (fig. 6) is constructed of V2A stainless steel and conforms to the World Meteorological Organization Hellmann-type standards. A 4-channel datalogger is built into this compact precipitation gage and can be interfaced with a laptop computer, HYDRAM II Reading Unit, or other external processing system. Data output is in serial digital RS232 form or as a pulse-output simulated tipping bucket. The gage can sense and record precipitation intensities up to about 2 inches per minute (50 mm). The accuracy of the gage is less than 0.0016 inch when collecting 0.39 inch of precipitation, whereas the sensitivity is 0.01 inch. The system uses a software compensation system to avoid variance due to temperature-related water-density differences. Several electronic filters are included to prevent wind-related artifacts in collected data. This gage does not require the use of a funnel; therefore, significant evaporation loss is prevented. While power is typically supplied by 12-volt direct current, 9- to 16-volt direct current can be used to operate the OTT PLUVIO gage (OTT Messtechnik GmbH and Company, written commun., 1999).

Figure 6. OTT PLUVIO rain gage.

NovaLynx National Weather Service Type Rain Gage

The NovaLynx stick gage (fig. 7) is an all-aluminum cylindrical gage that can measure up to 2 inches of precipitation in 0.01-inch increments in an inner chamber referred to as the receiver. The accuracy of the gage is 0.5 percent of full scale whereas the sensitivity is 0.01 inch. Excess rainfall overflows the inner chamber into the outer chamber where it can be measured after the quantity in the receiver has been measured and removed. The dipstick is marked in both English and metric equivalents. During the winter months, the receiver and funnel are removed so snowfall can be measured directly. This mechanical gage does not require power from an external source (NovaLynx Corporation, 2002).

Figure 7. NovaLynx Model 260–2510 National Weather Service type rain gage.

METHODS OF STUDY

Two identical gages of each type were evaluated during a two-phase testing program. Phase I involved testing of the gages indoors, in a controlled laboratory environment (68° to 75° F). Phase II involved testing the gages in an outdoor environment.

Phase I Testing

Phase I testing was performed indoors to test the accuracy and precision of the gages. With the exception of the Belfort 3200 gages, which were floor mounted, all of the gages were mounted on sturdy, level, 3- by 5-foot metal tables in the equipment-testing laboratory at the HIF. Before testing, each gage was calibrated according to the procedures and standards provided by the gage manufacturer. The pedestals and collection buckets for all of the gages also were leveled. Calibration coefficients were entered into a Campbell data-logger program as required for the various gages. The diameter of each inlet orifice was precisely measured and a gram/inch standard was determined for each gage by using manufacturer-specified milliliters of water equivalent to 1 inch of precipitation. These amounts were converted to temperature-compensated grams of water. Tap water was precisely weighed using a National Institute of Standards and Technology (NIST) traceable calibrated gram scale before being added to each gage. An individual test involved adding simulated rainfall amounts in small increments that produced cumulative depths from 0.01 inch to as much as 12 inches, unless the maximum full-scale capacity of the gage was less than 12 inches, in which case the gage was tested to its standard full-scale capacity. Twelve inches of rainfall is the capacity of the current NADP rain gage and is sufficient for the weekly or daily NADP monitoring networks; capacities of some of the gages in excess of 12 inches were therefore not evaluated. All of the gages that were tested are continuous recording gages, except the Novalynx stick gages. Unlike the other gages that were tested by adding small increments that produced cumulative depths, the NovaLynx stick gages were tested with separate amounts of water decanted for each depth. The 2-inch-capacity tube was completely dried before each depth was decanted. The equivalent of 2 inches of rainfall was decanted into the NovaLynx stick gage, the largest volume for an individual test. While the NovaLynx stick gage has an overflow chamber that provides a total gage capacity of 20 inches of precipitation, the inner chamber and stick are only calibrated to 2 inches of precipitation. Adding amounts of simulated rainfall greater than the 2-inch capacity of the inner chamber would have only tested the ability of the technician to pour water carefully and therefore was not done. The largest volume decanted at one time for the other gages was equivalent to 1 inch of rainfall. Each volume was decanted gradually and the gage reading allowed to stabilize before recording the gage response.

For the Geonor T–200, Belfort 5–780, Belfort 3200, and ETI Noah II, the cumulative depths of simulated rainfall applied were 0.00, 0.01, 0.05, 0.10, 0.30, 0.50, 0.75, 1.0, 2.00, 3.00, 4.00, 5.00, 6.00, 7.00, 8.00, 9.00, 10.00, 11.00, and 12.00 inches. For the OTT PLUVIO, the same cumulative depths were applied to 8.00 inches (effective gage capacity due to software problems). The cumulative depths applied for the NovaLynx stick gage were 0.00, 0.01, 0.05, 0.10, 0.25, 0.75, 1.05, 1.58, and 2.00 inches. Tap water in the depths specified was applied for a minimum of 20 times to each gage, which produced a minimum of 20 individual tests (except for the NovaLynx stick gages, which were each tested 19 times). Human error resulted in 28 missing values for the Belfort 5–780-2 gage; in four individual tests of this unit, the first seven depth values were inadvertently omitted. Minor technical problems or human errors resulted in a total of one or two missing values for three other gages over the entire course of Phase I testing.

As a check, an electrical depth gage was used to verify the amount of water present in each gage after the addition of water. As the specified amounts of water were added to each gage, the water was weighed twice, first with a Mettler PE 3600 digital scale and then with a Mettler P3N scale. The Mettler PE 3600 digital scale is accurate to $\pm$ 0.1 g. A Mettler P3N scale, accurate to $\pm$ 0.3 g at a full scale of 3,000 g, was used to double check the amounts of water added to each gage. Both scales were calibrated according to NIST standards before and after Phase I testing.

A Campbell Scientific CR10X datalogger and an RS232 electrical interface were used to collect and store data from the Belfort 3200, ETI Noah II, and Geonor T–200 on a laptop computer as a text file. Data from the OTT PLUVIO gage were obtained from a built-in HYDRAM II datalogger and an RS232 interface and were stored on a laptop computer as a text file.

Phase II Testing

Phase II testing was performed outdoors on the grounds of the HIF in Bay St. Louis, Miss., to test accuracy and precision. The rain gages were installed in a 60- by 60-ft array at the field-test site by using NADP site-location protocols for guidance. The gages were located near the northwest corner of the HIF, with the orifices placed the same distance above the ground and leveled (fig. 8). If a gage is not level, it will not accurately record the precipitation. A standard carpenter's level was used to check the gages' front-to-back and side-to-side levelness, and any adjustments also were checked with the carpenter's level. The gages are not within a 45-degree angle of any obstruction (fig. 9).

Gages were securely attached to their stands and firmly fixed in place. It was economically impractical to mount the gages on cement slabs. Platforms of different heights were constructed of treated 2 by 4’s and ¾-inch plywood to ensure that the orifices of all gages were at the same level. Gage stability was an important consideration; each gage was bolted to its platform, and sandbags were used on the platform cross members to ensure a solid base. Test data were obtained and the level of each gage observed while its platform was bumped repeatedly. It took a forceful kick to disturb the level or change the data by 0.01 inch. Placement of sandbags and gages was checked and adjusted after this test. Stability and levelness were also checked approximately every 2 weeks during the test. Adjustment was not needed over the course of the study for any of the gages.

Figure 8. Rain gages deployed for Phase II testing at the U.S. Geological Survey Hydrologic Instrumentation Facility in Bay St. Louis, Mississippi.

All gages were recalibrated according to the appropriate manufacturer specifications. Although some NADP sites shield their rain gages, it is not an NADP requirement to do so. To minimize the physical distance between gages, Alter-type shields were not used during Phase II testing. Legates and DeLiberty (1993) reported that the systematic undercatch biases in gage-measured summer precipitation in the region of Mississippi where Phase II testing was completed was between 4 and 5 percent. Systematic undercatch biases in gage-measured summer precipitation can be expected to vary slightly, depending on the diameter of the rain-gage opening (Golubev and others, 1992). Comparisons during Phase II testing were between gages and gage types; differences in gage performance related to differences in liquid precipitation undercatch are believed to have been minimal.

Each week HIF personnel retrieved the data for each test gage. Electronic data retrieval was performed as described in Phase I testing with the following exceptions: (1) The Belfort 5–780 paper charts were changed weekly and sent to the lead investigator for interpretation, and (2) visual inspection of precipitation depth for the NovaLynx stick gage was done daily by HIF personnel and manually recorded.

For the purpose of this study, evaluation of gage performance was determined on the basis of multiple test results. An accurate gage would return results identical to the various amounts of water applied during Phase I testing. During Phase II testing, the amounts of water recorded by the NovaLynx stick gages were used as the reference with which the other gages were compared. Precision would be indicated by repeatability or duplication of results between gages of the same type, in both phases of testing. Results also were evaluated using nonparametric statistical tests to compare gages of the same type and to compare gage performance against a known amount of applied simulated rainfall (Phase I) and against the results obtained from a reference gage (Phase I and II).

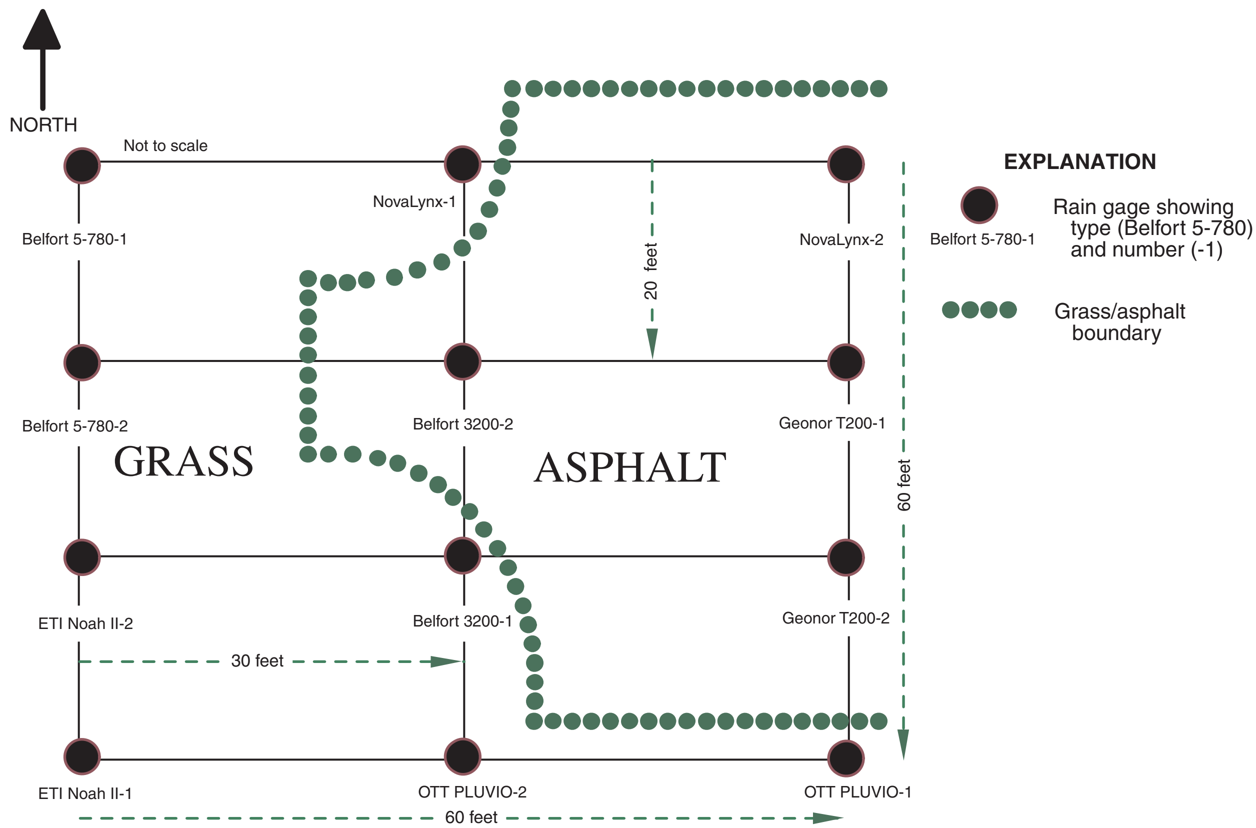


Figure 9. Diagrammatic plan view of the rain gages deployed for Phase II testing at the U.S. Geological Survey Hydrologic Instrumentation Facility in Bay St. Louis, Mississippi. Drawing is not to scale.

RESULTS OF EVALUATION

Rain-gage responses were arranged in chronological order and graphed for visual inspection to determine unusual or unexpected patterns or trends. While HIF personnel attempted to test the OTT PLUVIO gages to their 10-inch capacity in Phase I, a software algorithm problem limited meaningful readings to 8 inches. For amounts of applied water over 8 inches, results were omitted to remove the effects of the algorithm problem. The Belfort 3200 gages had technical problems that were difficult for HIF personnel to correct during both Phase I and Phase II testing.

Phase I Test Results

Phase I data were collected in December 1998 and January 1999. Table 2 shows selected statistics including the median, lower, and upper quartiles for the differences in inches between the measured and applied simulated rainfall during Phase I testing. In individual tests, large differences between the measured and applied amounts of simulated rainfall were occasionally observed. Differences between measured and applied simulated rainfall ranged from –0.601 to +0.270 inch for all gages tested. The difference of –0.601 inch was observed in a test where 11 inches of simulated rainfall was applied to the Geonor T–200–1 gage, which equates to a difference of –5.46 percent. The difference of +0.270 inch occurred in a test using the Belfort 5–780–2 gage when 5 inches of simulated rainfall was applied, which is the equivalent of a +5.40 percent difference (table 2).

The median difference, in inches, between the measured and applied simulated rainfall was 0.000 for all gages except the Belfort 3200 and the Geonor T–200. The median differences between the measured and applied simulated rainfall were –0.042 and –0.024 inch for the two Geonor T–200 gages and –0.110 and –0.024 inch for the two Belfort 3200 gages. Gages where both the median difference

Table 2. Selected statistics for the differences, in inches, between the measured and applied simulated rainfall during Phase I testing

[All units in inches; N, number of samples; Q1, the lower quartile in data distribution; Q3, the upper quartile in data distribution]

Gage	N	Missing values	Minimum	Maximum	Median	Quartile		Interquartile range
						Q1	Q3	
Belfort 3200–1	380	0	–0.412	0.062	–0.110	–0.228	–0.015	0.213
Belfort 3200–2	380	0	–0.299	0.093	–0.024	–0.085	0.002	0.087
Belfort 5780–1	380	0	–0.060	0.090	0.000	–0.010	0.000	0.010
Belfort 5780–2	352	28	–0.110	0.270	0.000	–0.020	0.000	0.020
Geonor T–200–1	380	0	–0.601	0.185	–0.042	–0.116	–0.008	0.108
Geonor T–200–2	380	0	–0.230	0.077	–0.024	–0.077	–0.001	0.076
ETI Noah II–1	379	1	–0.020	0.100	0.000	–0.001	0.000	0.001
ETI Noah II–2	378	2	–0.450	0.020	0.000	–0.001	0.000	0.001
OTT PLUVIO–1	300	0	–0.020	0.000	0.000	–0.010	0.000	0.010
OTT PLUVIO–2	299	1	–0.030	0.000	0.000	–0.010	0.000	0.010
NovaLynx NWS Type Gage–1	171	0	–0.090	0.040	0.000	0.000	0.000	0.000
NovaLynx NWS Type Gage–2	171	0	–0.090	0.040	0.000	0.000	0.000	0.000

between the measured and applied simulated rainfall and the interquartile range of all of their measured minus applied simulated rainfall differences were small (≤0.01inch) were judged to have performed very well in Phase I testing. The median and interquartile-range values were 0.01 inch or less for each of the ETI Noah II gages, OTT PLUVIO gages, and NovaLynx stick gages. Median and interquartile-range values approximating 0.00 inches indicate that gages produce unbiased results and results with low variability, respectively, under controlled (laboratory testing) conditions. The performance of the Geonor T–200 and Belfort 3200 gages was affected by technical problems during Phase I testing.

In figure 10, the median, spread, skewness, and presence or absence of outlying values for Phase I testing are depicted in boxplots (Helsel and Hirsch, 1992). The differences for all depths of applied simulated rainfall for the ETI Noah II, OTT PLUVIO, and NovaLynx stick gages were small, which indicates that the ETI Noah II, OTT PLUVIO, and NovaLynx stick gages consistently recorded depths more commensurate with the amounts of applied simulated rainfall in Phase I testing than the Geonor T–200, Belfort 5–780, and Belfort 3200 gages. Results of a Friedman test also showed that there were statistically significant differences ($\alpha = 0.01$) between all gages of different makes and models in the Phase I testing.

The precision of the results obtained from each pair of identical gages was evaluated to determine if there were statistically significant differences between paired gages of identical type. The null hypothesis that there was no difference between the paired gages of identical type was tested against the alternative hypothesis that there was a statistically significant difference between the paired gages of the same type. Results of the Wilcoxon signed-rank test indicated statistically significant differences between the Geonor T–200 paired gages ($\alpha = 0.01$, $p = 0.0075$) and the Belfort 3200 paired gages ($\alpha = 0.01$, $p = 0.0003$). The paired differences between all other gages of the same type were not statistically significant, which indicates good precision.

For each rain gage, differences between the measured and applied simulated rainfall were evaluated for each depth based on the individual tests. Cumulative absolute differences between the measured and applied amounts of water are shown as bar graphs (figs. 11 and 12) and listed numerically (table 3). A missing bar for a particular gage indicates a cumulative absolute difference of 0.00; this implies that the differences between the measured and applied simulated rainfall were exactly 0.00 for all of the tests for that gage at the indicated depth. Cumulative absolute differences of 0.00 at each depth applied were observed for at least one and, in many cases, for both of the OTT PLUVIO gages over the range of 0.01 to

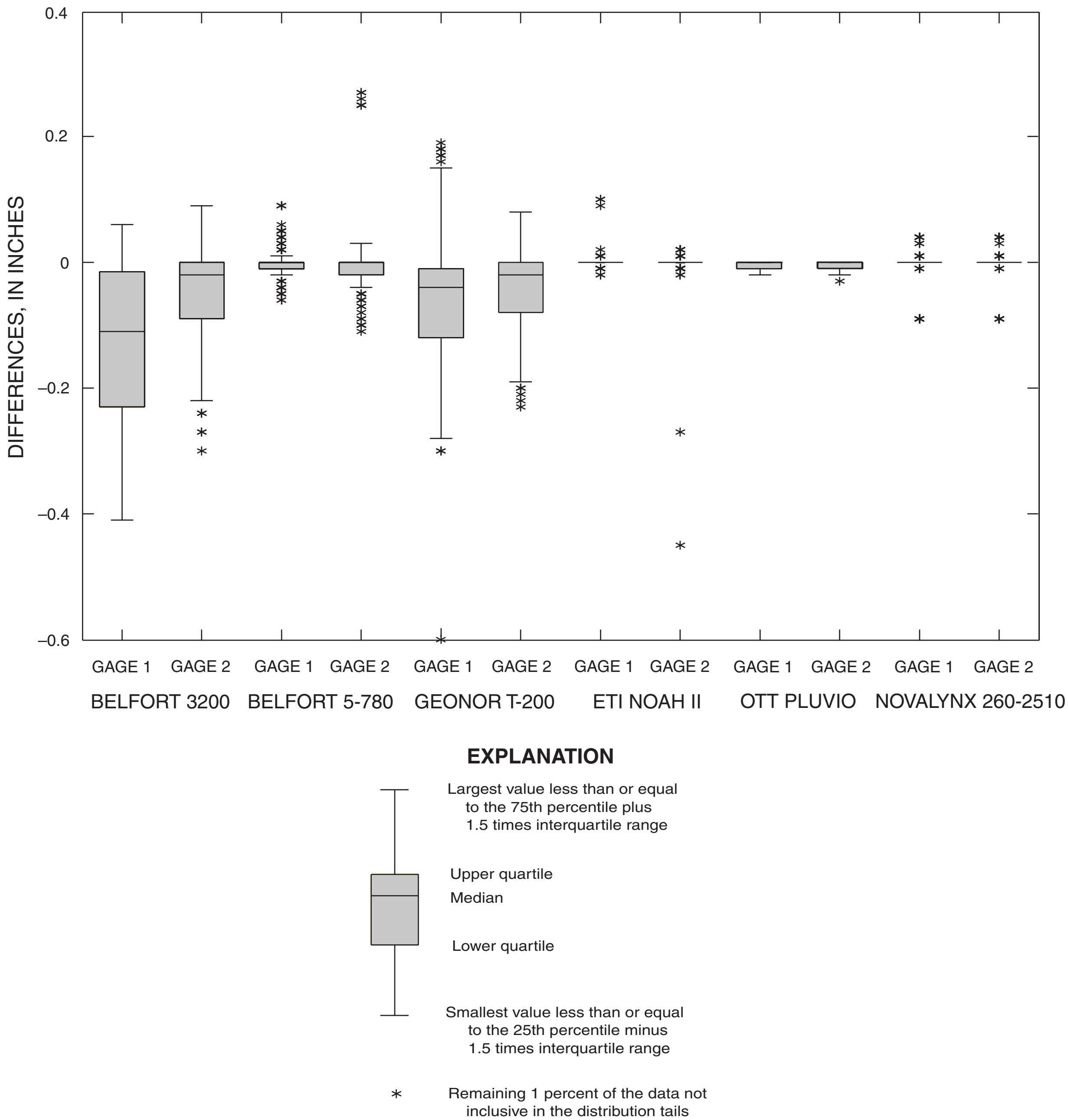
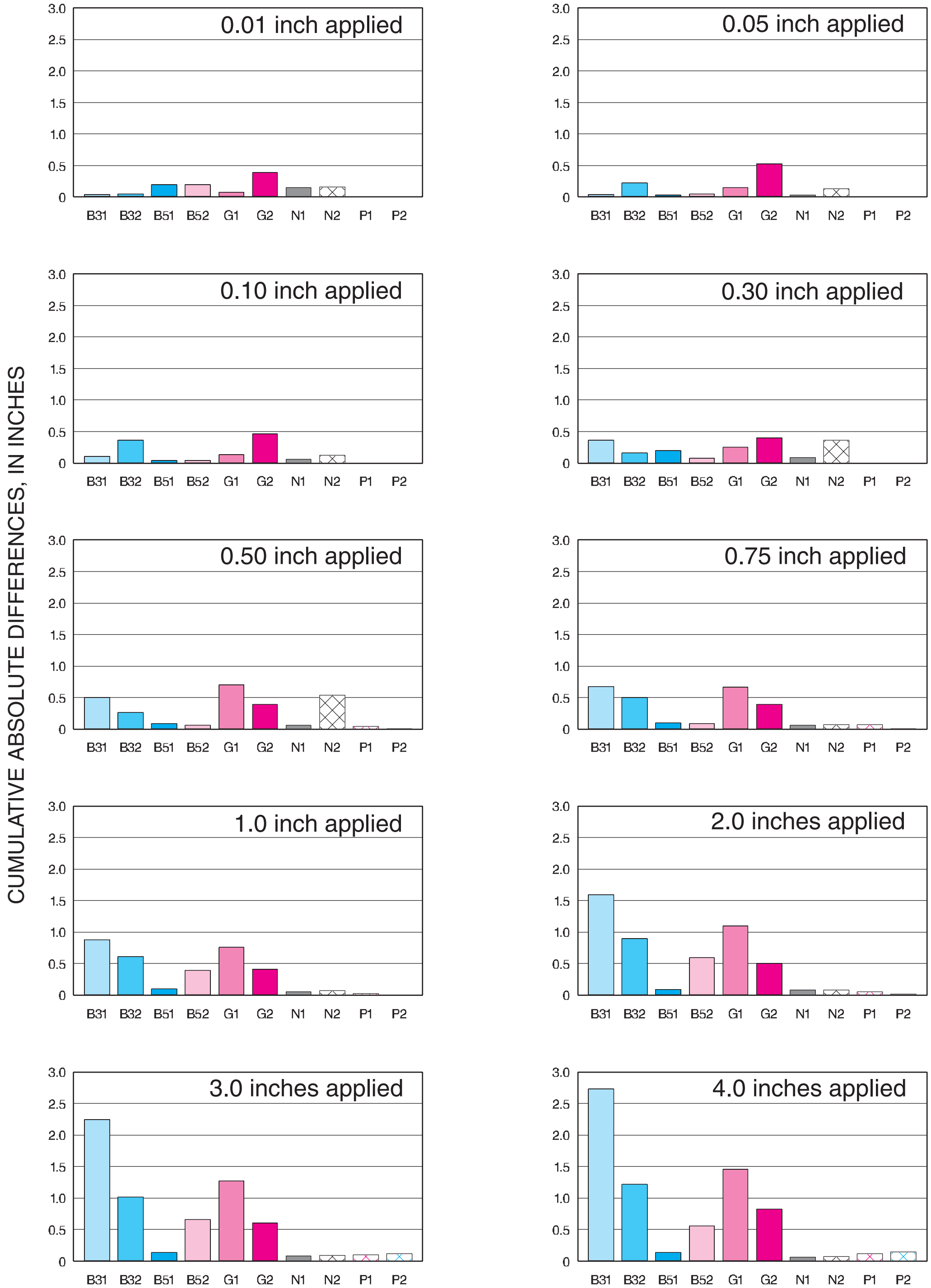


Figure 10. Differences, in inches, between the amount of simulated rainfall measured and the amount of simulated rainfall applied in Phase I testing (0.01 inch applied to 4.0 inches applied).

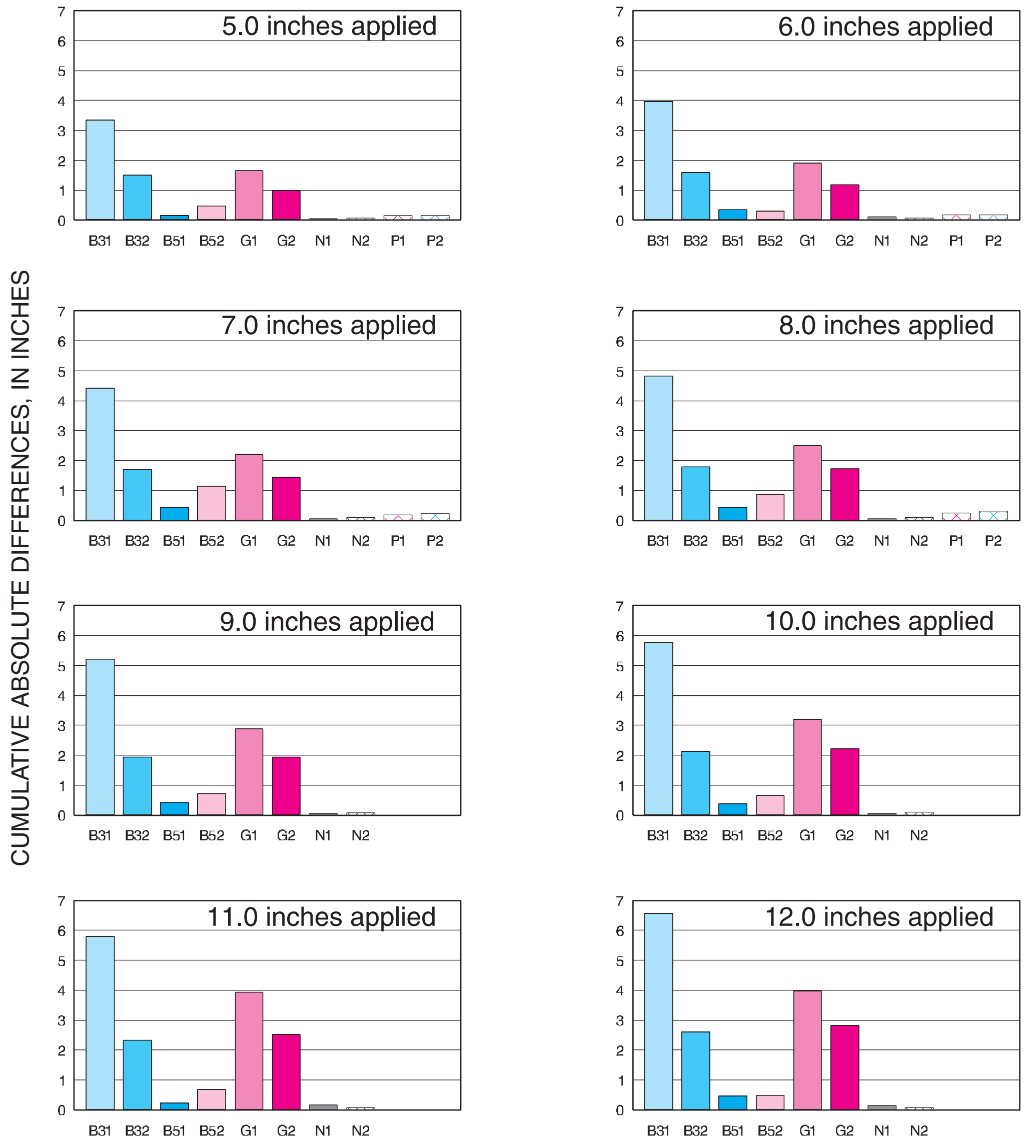
1.0 inch of applied simulated rainfall. This analysis is useful for determining which depths of applied simulated rainfall are accurately determined by individual gages. The cumulative asbsolute difference for each gage for depths of simulated rainfall of 3.0 and 8.0 inches indicates the relative accuracy of the gages tested. The gages are listed in table 3 in order of decreasing accuracy for the 3-inch and 8-inch simulated rainfall depths. For example, the Belfort 3200 gages, which did not perform well on an overall basis in Phase I testing accurately measured 0.01 and 0.05 inch of simulated rainfall and compared favorably with most gages for up to a depth of 1.0 inch of simulated rainfall. Between 2 and 12 inches of applied simulated rainfall, the cumulative absolute differences observed with the Belfort 3200–1 incrementally increased at a faster rate than the cumulative absolute differences observed with other gage types tested. The Belfort 3200–2 performed better than the Belfort 3200–1 over a range of 2 to 12 inches of applied rainfall. Figures 11 and 12 and table 3 show that the performance of the ETI Noah II gages was similar to



EXPLANATION

B31 = Belfort 3200-1 B32 = Belfort 3200-2
B51 = Belfort 5-780-1 B52 = Belfort 5-780-2
G1 = Geonor T-200-1 G2 = Geonor T-200-2
N1 = ETI Noah II-1 N2 = ETI Noah II-2
P1 = OTT PLUVIO-1 P2 = OTT PLUVIO-2

Figure 11. Cumulative absolute differences, in inches, between the amount of simulated rainfall measured and the amount of simulated rainfall applied in the Phase I testing (0.01 inch applied to 4.0 inches applied).



EXPLANATION

B31 = Belfort 3200-1	B32 = Belfort 3200-2
B51 = Belfort 5-780-1	B52 = Belfort 5-780-2
G1 = Geonor T-200-1	G2 = Geonor T-200-2
N1 = ETI Noah II-1	N2 = ETI Noah II-2
P1 = OTT PLUVIO-1	P2 = OTT PLUVIO-2

Figure 12. Cumulative absolute differences, in inches, between the amount of simulated rainfall measured and the amount of simulated rainfall applied in the Phase I testing (5.0 inches applied to 12.0 inches applied).

the performance of the OTT PLUVIO over much of the range of applied simulated rainfall, particularly for depths over the range of 0.75 to 8.0 inches. For extremely low amounts of applied simulated rainfall, 0.01 to 0.50 inch, the OTT PLUVIO cumulative absolute differences were smaller than those for other gages tested. On an overall basis, the OTT PLUVIO gages, when compared to results for other gages, had the lowest cumulative absolute differences for up to 2.0 inches of applied simulated rainfall. For 4 to 8 inches of simulated rainfall, the ETI Noah II gages had the lowest cumulative absolute differences, followed by the OTT PLUVIO gages. It also is worth noting that the gage currently in use throughout the NADP, the Belfort 5–780, did relatively well in Phase I testing, performing better than the Geonor T–200

Table 3. Cumulative absolute differences between the amount of simulated rainfall measured by each gage and the amount of simulated rainfall added to each gage during Phase I testing

Gage	Cumulative absolute difference, in inches													
Belfort 3200–1	0.04	0.04	0.11	0.36	0.50	0.68	0.88	1.59	2.25	2.73	3.35	3.96	4.41	4.82
Belfort 3200–2	0.05	0.22	0.36	0.16	0.26	0.50	0.61	0.90	1.02	1.22	1.51	1.60	1.70	1.79
Belfort 5780–1	0.20	0.03	0.04	0.20	0.09	0.10	0.10	0.09	0.14	0.14	0.16	0.36	0.44	0.45
Belfort 5780–2	0.20	0.05	0.04	0.08	0.06	0.09	0.39	0.59	0.66	0.56	0.47	0.30	1.14	0.86
Geonor T200–1	0.08	0.15	0.13	0.25	0.70	0.67	0.76	1.10	1.27	1.46	1.65	1.91	2.19	2.49
Geonor T200–2	0.39	0.53	0.46	0.40	0.39	0.39	0.41	0.50	0.60	0.82	1.00	1.19	1.45	1.73
ETI Noah II–1	0.15	0.03	0.06	0.09	0.06	0.06	0.05	0.08	0.08	0.06	0.06	0.12	0.06	0.05
ETI Noah II–2	0.16	0.13	0.12	0.36	0.54	0.07	0.07	0.08	0.09	0.07	0.08	0.08	0.10	0.10
OTT PLUVIO–1	0.00	0.00	0.00	0.00	0.04	0.07	0.03	0.05	0.10	0.12	0.16	0.18	0.18	0.24
OTT PLUVIO–1	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.02	0.12	0.15	0.16	0.18	0.22	0.31
	Depth of simulated rainfall added, in inches													
	0.01	0.05	0.10	0.30	0.50	0.75	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00

and Belfort 3200 gages in the Phase I cumulative absolute difference analysis but not as well as the OTT PLUVIO or ETI Noah II gages. This can be discerned by comparing the Belfort 5–780 cumulative absolute differences in figures 11 and 12 and table 3 with the cumulative absolute differences for the other gages that were tested.

Figure 13 shows the median difference, in inches, between the measured and applied amounts of rainfall for the individual tests at each gage. The

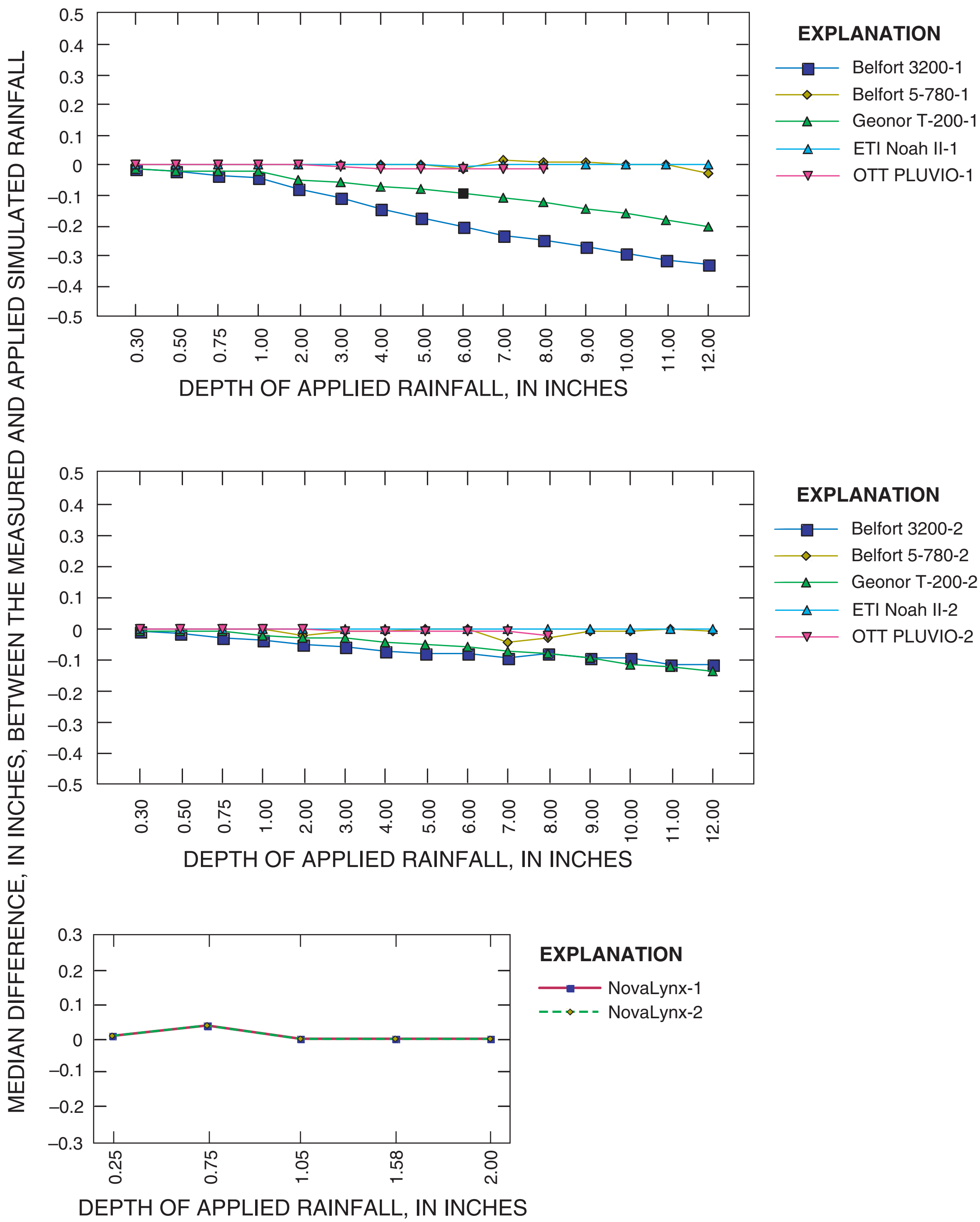


Figure 13. Median differences, in inches, between the measured and the applied simulated rainfall in Phase I testing.

median differences approximated 0.00 at all depths for the ETI Noah II and OTT PLUVIO gages, as well as for the NovaLynx stick gage. Median differences for the NovaLynx stick gage were plotted separately because of the unique depths of simulated rainfall applied for this gage and the NovaLynx stick gages only were tested to their inner-chamber capacity of 2.00 inches of rainfall. The median differences between the measured and applied simulated rainfall as percentages of the applied simulated rainfall are shown in figure 14. For all gages, the median percent difference at each depth applied was between –5 and +5 percent. For both units of the ETI Noah II and the OTT PLUVIO gages, the median percent differences were within the narrow range of –0.5 to +0.5 percent.

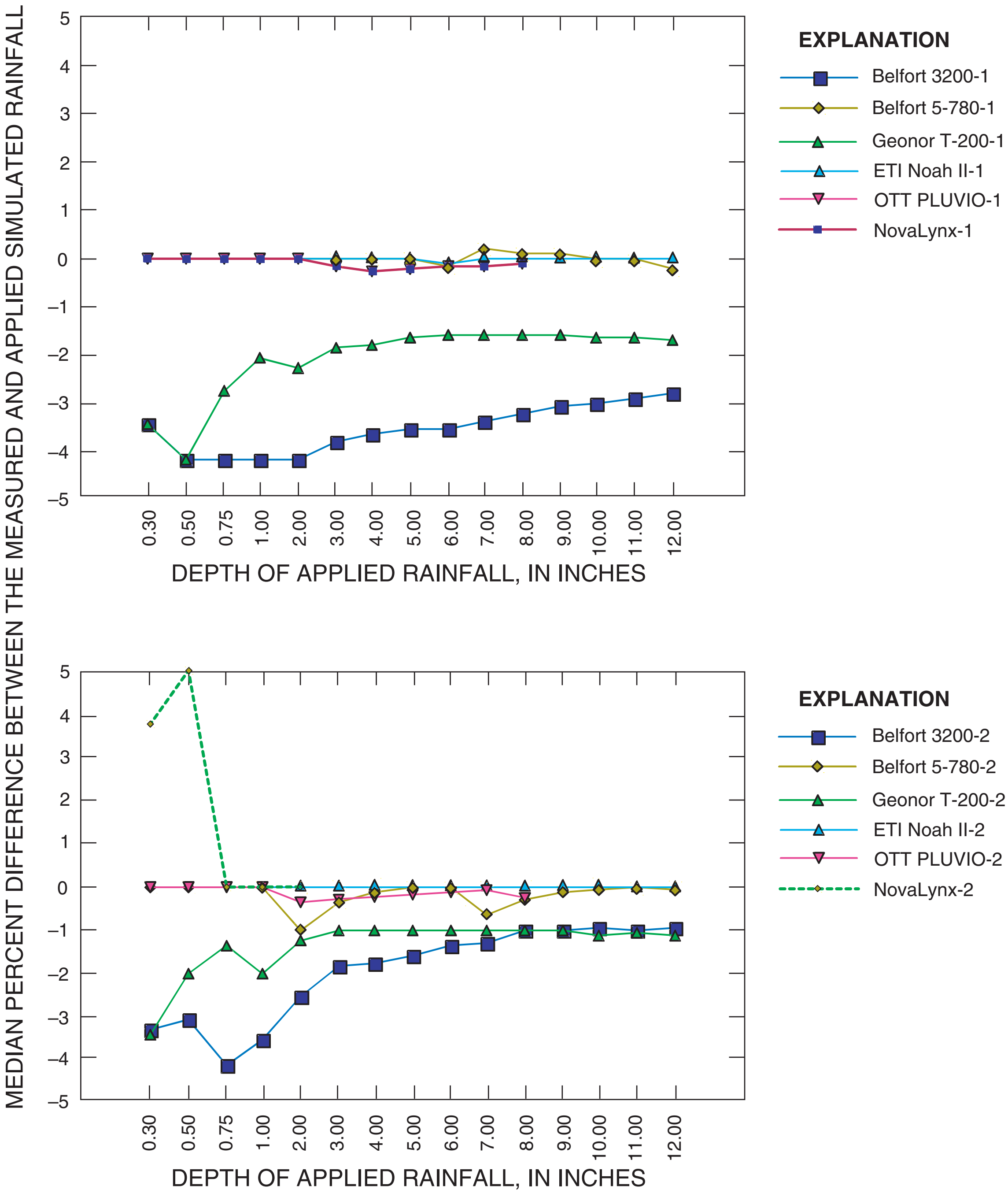


Figure 14. Median percent difference between the measured and applied simulated rainfall during Phase I testing of all gages.

Phase II Test Results

Following Phase I indoor bench testing, all of the gages were relocated outdoors at the HIF for the Phase II outdoor testing. Phase II data were collected during a 26-week testing period between January 8 and July 8, 1999. All precipitation during Phase II testing was in the form of rain. Except for the NovaLynx stick gages, which were read once a day, all other gages were set to record precipitation continuously. The Belfort 5–780 gages were equipped with paper recording charts and operated in a manner simulating their use by the NADP. Weekly totals for the Belfort 5–780 gages were determined from the paper charts. With the exception of the NovaLynx stick and Belfort 5–780 gages, the gages transmitted precipitation values every 15 minutes to dataloggers. Data periodically were downloaded from the dataloggers to a laptop computer by HIF personnel and printed out on paper. The original computer files were not available for this report, but USGS personnel entered the data into computer files from the paper copies.

The portions of the Phase II testing period for which data were obtained from each of the gages were different. The precipitation records for the OTT PLUVIO gages were complete for the entire 26 weeks of outdoor testing. Both Geonor T–200 gages operated for the 26 weeks of the study with the exception of a few brief periods due to battery failures, which fortuitously were at times when no precipitation was documented by the other gages in the study. The OTT PLUVIO gage was the most reliable gage in both Phase I and II testing. The gage operated trouble free over the duration of the study. The ETI Noah II–1 gage operated for 24 of the 26 weeks, and the ETI Noah II–2 gage operated for 16 of the 26 weeks. Mechanical and calibration difficulties with the Belfort 3200–1 and Belfort 3200–2 gages rendered them inoperable for 17 of 26 and 20 of 26 weeks of testing, respectively. The Belfort 5–780–1 and Belfort 5–780–2 gages operated for 25 of 26 and 23 of 26 weeks, respectively, and the periods of lost record for these gages were due to common operator errors with this gage (failure to rewind the clock, reink the pen tip, or change the chart). The two NovaLynx stick gages were read daily for the first 25 weeks of the 26-week study but were inadvertently not read by HIF personnel during week 26. It is worth noting that, for the gages with more than 3 weeks of lost record, the data loss was in sequential weeks beginning in week 1 of the Phase II testing. Once the gages that lost 3 or more weeks of record starting in week 1 of Phase II testing were repaired, they generally operated trouble free for the remainder of the study.

With the exception of the NovaLynx stick gages that were read daily, weekly totals were calculated for each gage by summing the continuous, 15-minute-increment precipitation values into daily totals, then adding the daily totals to obtain weekly totals. Care was taken to include only the upward movement in recorded values that was due to precipitation and to account accurately for shifts in the baseline for each gage—not a trivial task. Increases in a gage's baseline due to the addition of oil or water by the technicians were accounted for in the determination of daily precipitation totals. Decreases in the recorded values that occurred when water was lost from a gage due to evaporation or when liquid was removed from the gage by the technicians also was factored out of the precipitation totals.

Table 4 shows the weekly precipitation totals for each gage during Phase II testing; figures 15–16 depict weekly totals graphically. In 8 of the 26 weeks, weekly totals between 0.42 and 0.96 inch of precipitation were recorded by the NovaLynx stick reference gages. Differences in weekly totals among all of the gages in the study ranged from 0.06 to 0.18 inch when the weekly total precipitation measured by the reference gages was between 0.42 and 0.96 inch. In 6 of the 26 weeks, the amount of precipitation measured by the NovaLynx stick reference gages was between 1.43 and 3.18 inches, and all gages in the study measured total precipitation within 0.17 and 0.39 inch during these weeks. The maximum weekly difference among gages in a given week was 0.39 inch, which occurred during the wettest week of the study when 3.18 inches of precipitation was recorded by the reference gages.

Measurable precipitation (more than 0.01 inch) was recorded by one or more gages in 25 of the 26 weeks of testing. During Phase II testing, accurately measuring trace amounts of precipitation or weekly totals of no precipitation proved somewhat challenging. Three gages (Belfort 5–780, ETI Noah II, and OTT PLUVIO) each recorded 5 weeks with no precipitation, while the NovaLynx stick gage recorded 4 weeks with no precipitation and the Geonor T–200 gages recorded only 1 week with no precipitation.

Throughout Phase II testing, the Geonor T–200 weekly totals frequently were slightly higher than the weekly totals for the other technologically advanced

Table 4. Weekly precipitation totals, in inches, measured by each gage during Phase II testing

[--, no data]

Week	Dates	Belfort 3200 gage 1	Belfort 3200 gage 2	Belfort 5780 gage 1	Belfort 5780 gage 2	Geonor T–200 gage 1	Geonor T–200 gage 2	ETI Noah II gage 1	ETI Noah II gage 2	OTT PLUVIO gage 1	OTT PLUVIO gage 2	NovaLynx National Weather Service type gage 1	NovaLynx National Weather Service type gage 2	Maximum weekly total from all gages	Minimum weekly total from all gages	Maximum weekly difference among all gages
1	Jan 8–14	--	--	1.53	1.50	1.41	1.44	--	--	1.36	1.36	1.43	1.43	1.53	1.36	0.17
2	Jan 15–21	--	--	0.00	0.00	0.00	0.00	--	--	0.00	0.00	0.005	0.005	0.01	0.00	0.01
3	Jan 22–28	--	--	0.75	0.70	0.59	0.60	0.63	--	0.66	0.66	0.68	0.68	0.75	0.59	0.16
4	Jan 29– Feb 4	--	--	0.39	0.35	0.37	0.32	0.34	--	0.38	0.38	0.43	0.42	0.43	0.32	0.11
5	Feb 5 –11	--	--	0.05	0.05	0.06	0.06	0.03	--	0.05	0.05	0.07	0.06	0.07	0.03	0.04
6	Feb 12–18	--	--	0.90	0.83	0.87	0.84	0.84	--	0.85	0.88	0.93	0.90	0.93	0.83	0.10
7	Feb 19–25	--	--	0.11	--	0.13	0.16	0.09	--	0.11	0.11	0.12	0.12	0.16	0.09	0.07
8	Feb 26–March 4	--	--	0.73	0.72	0.74	0.75	0.70	--	0.71	0.71	0.88	0.86	0.88	0.70	0.18
9	March 5–11	--	--	0.45	.	0.44	0.46	0.42	--	0.43	0.42	0.48	0.47	0.48	0.42	0.06
10	March 12–18	--	--	2.88	2.86	2.88	2.87	2.84	--	2.87	2.85	3.01	3.01	3.01	2.84	0.17
11	March 19–25	--	--	0.00	0.00	0.07	0.11	0.00	0.00	0.00	0.00	0.02	0.01	0.11	0.00	0.11
12	March 26–April 1	--	--	--	--	2.35	2.34	2.32	2.26	2.33	2.34	2.48	2.48	2.48	2.26	0.22
13	April 2–8	--	--	0.04	0.04	0.01	0.04	0.00	0.00	0.01	0.00	0.03	0.03	0.04	0.00	0.04
14	April 9–15	--	--	0.00	0.00	0.04	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.04
15	April 16–22	--	--	0.00	0.00	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.05
16	April 23–29	--	--	0.00	0.00	0.10	0.08	0.04	0.05	0.08	0.08	0.05	0.04	0.10	0.00	0.10
17	April 30 –May 6	--	--	0.35	0.38	0.44	0.43	0.38	0.36	0.42	0.41	0.42	0.43	0.44	0.35	0.09
18	May 7–13	0.90	--	0.90	0.93	0.92	0.91	0.90	0.84	0.90	0.92	0.96	0.96	0.96	0.84	0.12
19	May 14–20	0.00	--	0.00	0.00	0.10	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.10
20	May 21–27	0.42	--	0.42	0.45	0.48	0.44	0.41	--	0.44	0.44	0.44	0.44	0.48	0.41	0.07
21	May 28–June 3	1.73	1.58	1.79	1.78	1.78	1.76	1.70	--	1.79	1.82	1.87	1.87	1.87	1.70	0.29
22	June 4–10	0.00	0.00	0.05	0.08	0.23	0.15	0.04	0.03	0.06	0.06	0.02	0.02	0.23	0.00	0.23
23	June 11–17	3.17	2.85	2.96	2.95	3.19	3.04	2.92	2.80	3.10	3.13	2.88	2.87	3.19	2.80	0.39
24	June 18–24	0.00	0.00	0.14	0.15	0.20	0.20	0.02	0.00	0.08	0.10	0.04	0.04	0.20	0.00	0.20
25	June 25–July 1	3.23	2.89	3.08	2.96	3.13	3.08	2.98	2.95	3.09	3.08	3.18	3.18	3.18	2.95	0.34
26	July 2–8	0.74	0.68	0.78	0.80	0.82	0.77	0.65	0.64	0.79	0.81	--	--	0.82	0.64	0.18

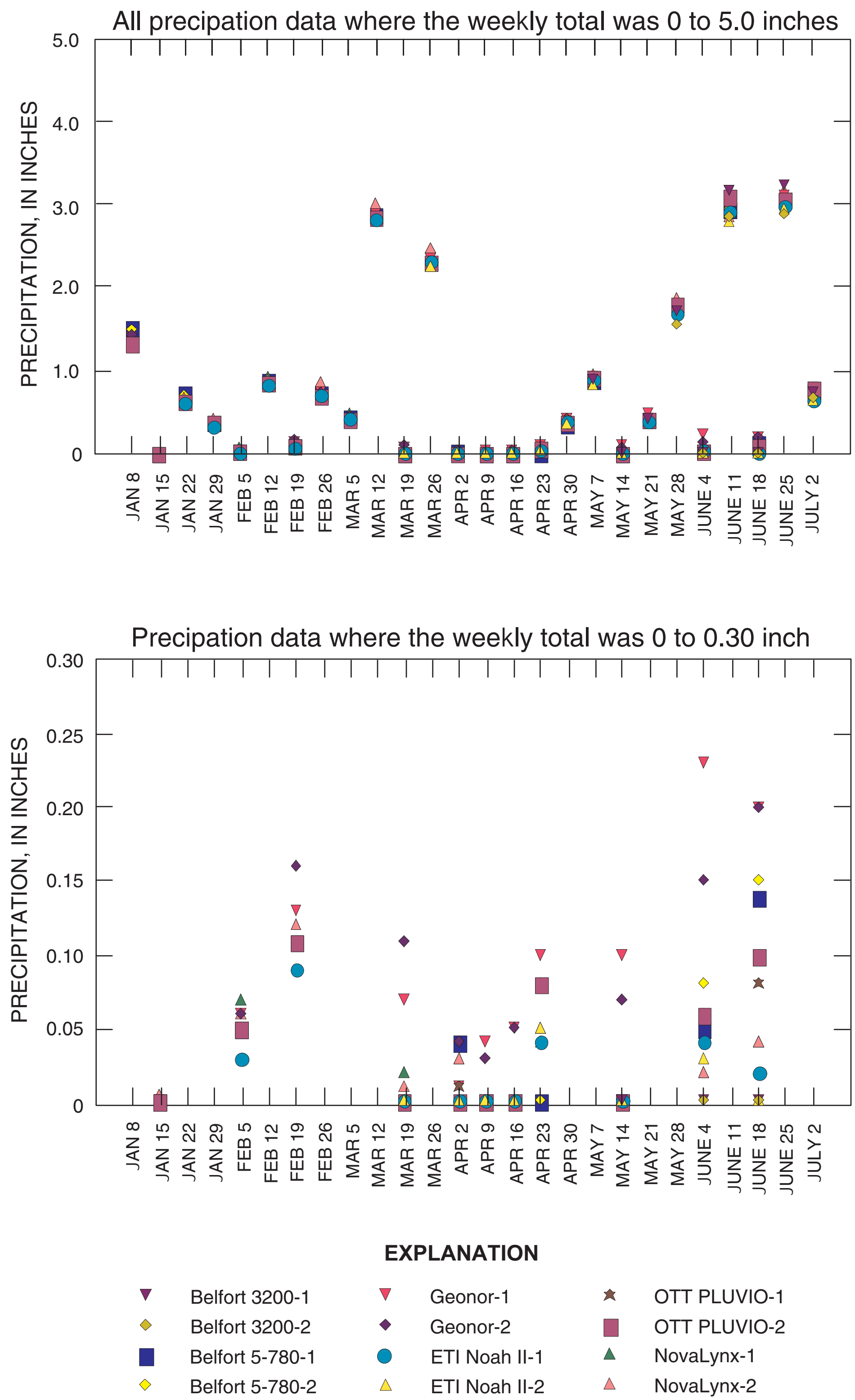


Figure 15. Weekly precipitation totals measured by each gage type during the 26-week Phase II testing period (0.00 to 0.30 inch and 0.00 to 5.00 inches of precipitation).

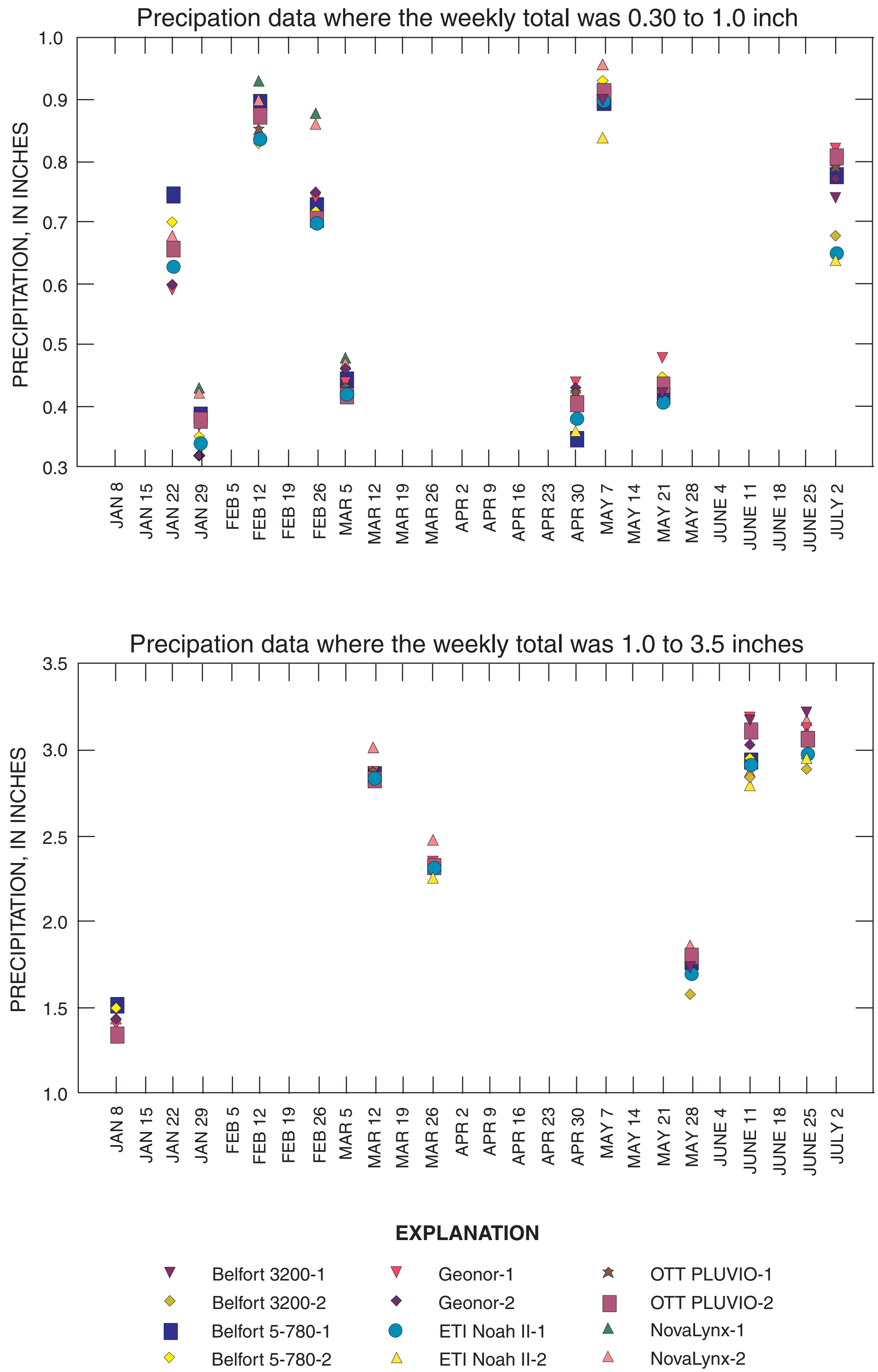


Figure 16. Weekly precipitation totals measured by each gage type during the 26-week Phase II testing period (0.30 to 1.0 inch and 1.0 to 3.5 inches of precipitation).

rain gages tested due to a few sporadic recordings of trace amounts of precipitation over the course of a week when other gages were recording no precipitation. Because manual observations were made once per day, it is not possible to know with absolute certainty if these sporadic trace values were errors or accurate recordings of precipitation. In 15 out of 20 weeks in which the Geonor T–200 and OTT PLUVIO gages both recorded measurable (0.01 inch or more) precipitation, the Geonor T–200 gages recorded a slightly higher weekly total than the OTT PLUVIO gages. The differences in weekly totals between the Geonor T–200 and OTT PLUVIO gages ranged from –0.07 to 0.13 inch, with a median difference of 0.02 inch. The Geonor T–200 gages also recorded slightly higher weekly totals in 15 out of 18 weeks in which the Geonor T–200 and ETI Noah II gages were both operational and recording measurable amounts of precipitation. The differences in weekly totals between the Geonor T–200 and ETI Noah II gages ranged from –0.04 to 0.26 inch, with a median difference of 0.05 inch. The Geonor T–200 weekly totals exceeded the Belfort 5–780 weekly totals in 12 of 17 weeks in which both gages recorded measurable precipitation. The differences in weekly totals between the Geonor T–200 and Belfort 5–780 gages ranged from –0.13 to 0.26 inch, with a median difference of 0.02 inch. Wind interference resulting in false positives and the temporary installation methods used for the study, which were less isolated from the effects of slight vibrations than recommended by the manufacturer, are believed to have contributed to the Geonor T–200 gages recording small amounts of precipitation when the ETI Noah II and the OTT PLUVIO gages were recording no precipitation. The ETI Noah II and the OTT PLUVIO gages came equipped with built-in filtering software that automatically removed signal noise resulting from wind and other sources of vibration, but the Geonor T–200 gage was not equipped with noise-filtering software.

The average weekly precipitation totals obtained from each of the gages under consideration as possible replacement gages for NADP use and from the two Belfort 5–780 gages included in the test to represent the current NADP gages were compared with the average weekly totals obtained from two NovaLynx stick gages to evaluate accuracy. Figure 17 depicts the difference, in inches, between the average weekly precipitation totals measured by each gage type and the average weekly total measured using two collo-

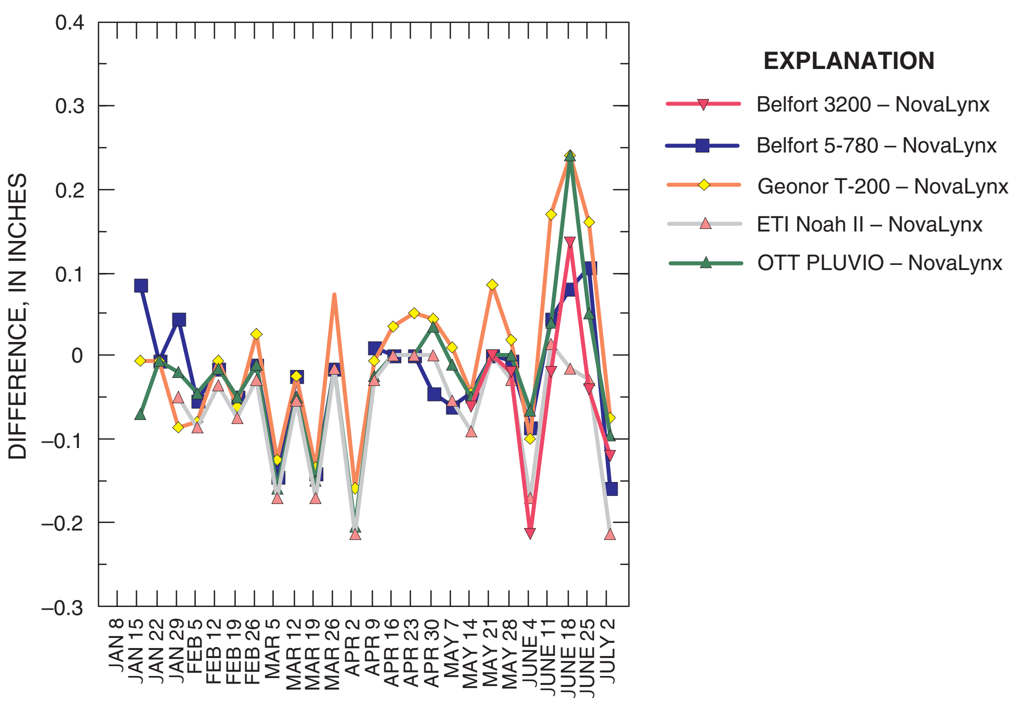


Figure 17. Difference, in inches, between the average weekly precipitation totals measured by each gage type and the average weekly precipitation total measured using two collocated NovaLynx Model 260–2510 National Weather Service type stick gages during Phase II testing.

cated NovaLynx stick gages. Over the entire time period of Phase II testing, the total absolute difference between average weekly totals measured with a particular gage type and the average weekly total measured with the NovaLynx stick reference gages ranged from 1.23 inches for the Belfort 5–780 to 1.83 inches for the Geonor T–200 gages (table 5). The total absolute difference is the sum of the weekly absolute differences between a particular gage type and the NovaLynx stick reference gages. The median absolute differences between a particular gage type and the NovaLynx stick reference gages for the 26 weeks of outdoor testing ranged from 0.04 inch for the OTT PLUVIO and the ETI Noah II gages to 0.06 inch for the Geonor T–200. Because the Belfort 3200 gages were inoperable for most of the Phase II testing, it is not meaningful to include the results from this gage type in a calculation of median or total absolute differences. Boxplots in figure 18 depict the overall distribution of differences between the average weekly precipitation measured by each gage and the average weekly precipitation total measured using two NovaLynx stick gages during Phase II testing. The median difference between average weekly precipitation totals calculated for each gage and the average weekly precipitation total for NovaLynx stick gages was negative with the exception of the Geonor T–200 gages. While the variability in paired Geonor T–200 minus NovaLynx stick gage differences was greater than the variability in paired comparisons between the other gages tested and the NovaLynx stick gage, the Wilcoxon signed-rank test results indicate no statistically significant differences between results for any of the rain gages and results for NovaLynx stick gages with the exception of the results for ETI Noah II gages ($p = 0.04$, $\alpha = 0.05$).

The results of the Phase II testing also were analyzed to determine the precision between gages of the same type (paired gages). Ideally, gages of the same type will record identical precipitation amounts. Table 6 lists summary statistics regarding the precision between two gages of the same type. The median relative differences of measured average weekly precipitation between gages of the same type ranged from 0.00 to 0.01 inch for all of the gages except the Belfort 3200, which had operational difficulties as mentioned previously. The total absolute differences between gages of the same type over the 26 weeks of the study were 0.12 inch for the NovaLynx stick gages, 0.22 inch for the OTT PLUVIO gages, 0.28 inch for the ETI Noah II gages, 0.51 inch for the Belfort 5–780 gages, and 0.75 inch for the Geonor T–200 gages, indicating good precision between measurements by the paired gages. Table 7 lists the results of the

Table 5. Differences in the average weekly precipitation total, in inches, for each gage type as compared to the average weekly total precipitation, in inches, determined from two NovaLynx National Weather Service type reference gages during Phase II testing

[All units in inches except for median relative percent difference and median absolute percent difference, which are in percent]

Gage average – Reference gage average	Belfort 3200 gage[1] – Reference gage[2]	Belfort 5–780 gage – Reference gage[2]	Geonor T–200 gage – Reference gage[2]	ETI Noah II gage[3] – Reference gage[2]	OTT PLUVIO gage – Reference gage[2]
Total relative difference	–0.34	–0.49	0.01	–1.52	–0.60
Median relative difference	–0.03	–0.01	–0.01	–0.04	–0.02
Total absolute difference[4]	0.61	1.23	1.83	1.55	1.41
Median absolute difference	0.05	0.05	0.06	0.04	0.04
Median relative percent difference	–5.40	–5.15	–4.59	–8.20	–4.90
Median absolute percent difference	5.47	10.64	8.02	8.43	8.27

[1]In 16 of the 25 weeks of the Phase II study, at least one of the Belfort 3200 gages was inoperational, making it difficult to compare its results to the other gages.

[2]NovaLynx National Weather Service type gage was used as the reference gage.

[3]At least one ETI Noah II gage was operational in 23 of the first 25 weeks of the Phase II study.

[4]Total absolute difference for gages for which there were not data for the entire study were extrapolated by multiplying by 1/fraction of the study where data were available.

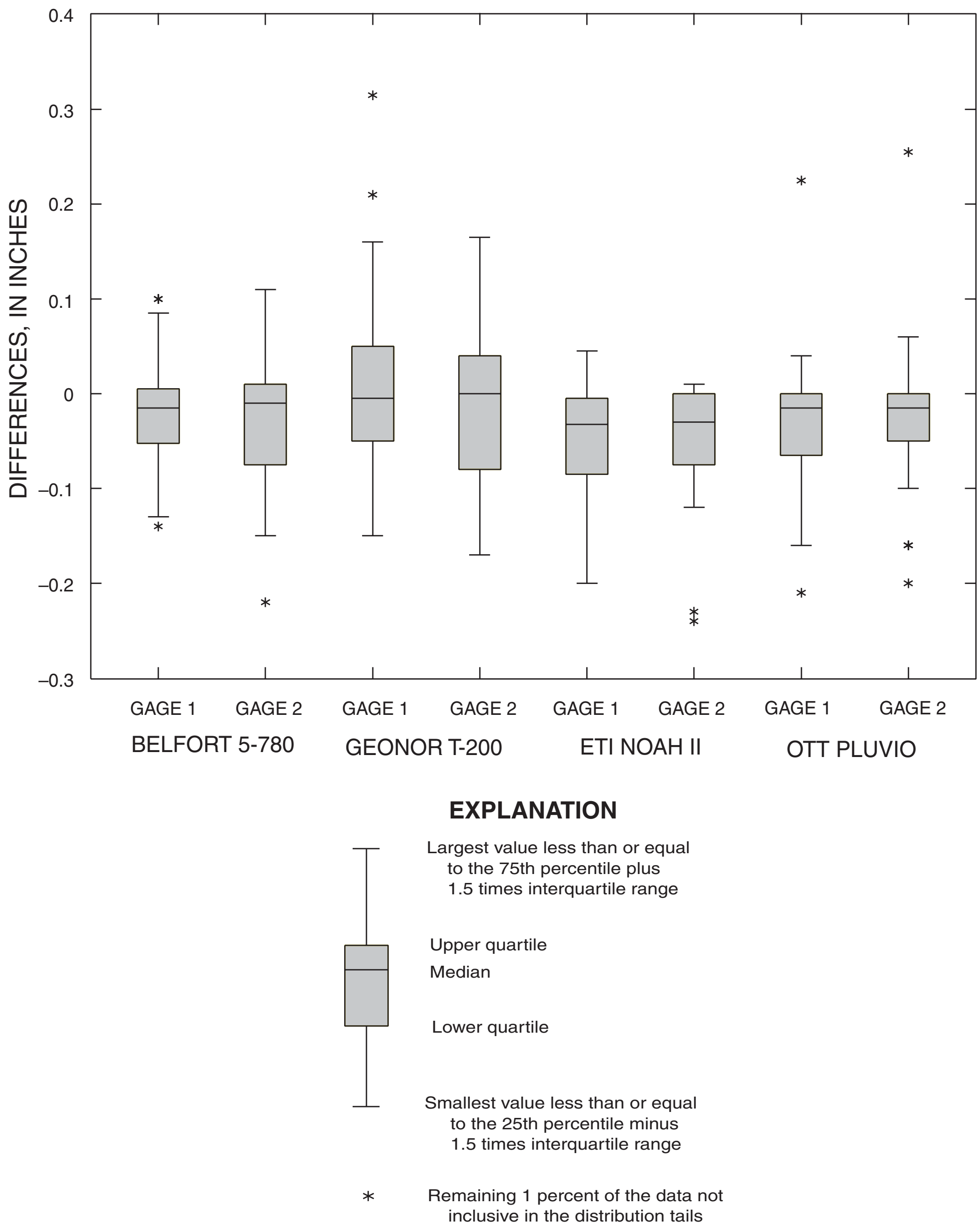


Figure 18. Differences, in inches, between the weekly precipitation total measured by each gage and the average weekly precipitation total measured by two NovaLynx Model 260–2510 National Weather Service type gages during Phase II testing.

Wilcoxon signed-rank test for the paired-gage comparisons. In all cases, the null hypothesis that there was no difference between the paired gages of the same type and model was not rejected, indicating that the small differences in measurements between gages of the same type were not statistically significant.

The OTT PLUVIO gage was the most reliable gage in Phase I and II testing. The gage operated trouble free over the duration of the study.

CONCLUSIONS

The OTT PLUVIO proved to be the most reliable gage in Phase I and II testing, operating trouble free over the duration of the study. The ETI Noah II gage and the OTT PLUVIO gage produced data of the highest accuracy and precision. The Geonor T–200 gage also performed reasonably well in Phase I and II testing and could be a suitable replacement gage for

Table 6. Phase II testing gage intracomparison—precision between two gages of the same type

[Units in inches except for the median relative percent difference and median absolute percent difference]

Paired gage statistics	[1]Belfort 3200 Gage1 – Gage2	Belfort 5–780 Gage1 – Gage2	Geonor T–200 Gage1 – Gage2	[2]ETI Noah II Gage1 – Gage2	OTT PLUVIO Gage1 – Gage2	NovaLynx NWS Gage1 – Gage2
Total relative difference	0.87	0.21	0.41	0.31	–0.10	0.10
Median relative difference	0.11	0.00	0.01	0.01	0.00	0.00
Total absolute difference	0.87	0.51	0.75	0.28	0.22	0.12
Median absolute difference	0.11	0.01	0.02	0.01	0.00	0.00
Median relative percent difference	8.76	0.00	0.00	1.28	0.00	0.00
Median absolute percent difference	8.76	0.70	1.89	1.88	0.00	0.00

[1]Both Belfort 3200 gages only operated 5 of first 25 weeks of the Phase II study.
[2]Both ETI Noah II gages only operated 13 of first 25 weeks of the Phase II study.

Table 7. Wilcoxon signed-rank test results for paired gage comparisons during Phase II testing

[>, greater than; <, less than; =, equals; Probability (Prob.) number of outcomes in the event divided by the total number of outcomes in the sample space; α, the maximum probability of making a Type I error (rejecting a null hypothesis when it is true); Z, test statistic for the Wilcoxon signed-rank test]

Gage type	Gage 1 > Gage 2	Gage 1 < Gage 2	Gage 1 = Gage 2	Weekly results from one or both gages missing	Z	Prob. > |Z|	Determined to be biased ($\alpha = .01$)
Belfort 3200	4	0	2	20	–0.021	0.983	NO
Belfort 5–780	9	6	8	3	0.078	0.938	NO
Geonor T–200	15	7	4	0	0.018	0.985	NO
ETI Noah II	8	1	5	12	0.101	0.920	NO
OTT PLUVIO	5	7	14	0	–0.904	0.366	NO
NovaLynx NWS type stick gage	8	1	16	1	0.000	1.000	NO

the NADP if the technical problems that kept this gage from performing to the same level as the OTT PLUVIO and ETI Noah II gages can be overcome. The Belfort 3200 gages could also be considered as possible replacement gages if redesigning the gage can correct the technical problems that affected its performance in Phase I and II testing. A winter study to evaluate gage performance under adverse snow and ice conditions must be conducted before any final recommendation is made to NADP personnel.

While evaluating the performance of various technologically advanced rain gages, it is important to keep in mind that wind-induced gage undercatch of precipitation, among other known systematic errors, is the greatest source of bias in precipitation observation (Yang and others, 1999). Currently, rain-gage shielding is optional at NADP sites. More NADP sites lack shielding than currently have shielding for their rain gages. The combination of precipitation records from shielded and unshielded gages can result in inhomogeneous precipitation time series and can lead to incorrect spatial interpretations (Yang and others, 1999). Because time-series analyses on atmospheric deposition data collected by the NADP currently rely on shielded and unshielded rain gages, it can be inferred that the same type of errors in deposition time-series analyses using NADP precipitation data may be occurring, as observed by Yang and others (1999) and many other investigators.

SUMMARY

The U.S. Geological Survey evaluated the performance of four technologically advanced rain gages as possible replacement gages for the current mechanical gage (Belfort 5–780) in use at all NADP precipitation monitoring sites. The gage models evaluated were the Belfort 3200, Geonor T–200, ETI Noah II, and the OTT PLUVIO. The NovaLynx National Weather Service type stick gage also was included in the study as a reference gage. The rain gages were tested for accuracy, precision, and reliability in a two-phase study that was done by the U.S. Geological Survey Hydrologic Instrumentation Facility in Bay St. Louis, Mississippi. In the first phase, gages were bench tested in a laboratory setting with known amounts of simulated rainfall applied in small increments to the full capacity of each gage or up to 12 inches, depending on the gage type. In Phase I tests, the median difference between the measured and applied simulated rainfall was 0.000 inches for all gages except the Belfort 3200 and the Geonor T–200. The median differences between the measured and applied simulated rainfall were –0.042 inch and –0.024 inch for the two Geonor T–200 gages. For the two Belfort 3200 gages, median differences of –0.110 inch and –0.024 inch were measured.

The reproducibility of the results obtained from each pair of identical gages was evaluated to determine if there were statistically significant differences between paired gages of identical type. Results of the Wilcoxon signed-rank test indicated statistically significant differences between the Geonor T–200 paired gages ($p = 0.0075$) and the Belfort 3200 paired gages ($p = 0.0003$). The paired differences between all other gages of the same type were not statistically significant.

Differences between the measured and applied simulated rainfall were evaluated for each depth that was tested. Cumulative absolute differences of 0.00 inches at each depth applied were observed for at least one and in many cases for both of the OTT PLUVIO gages over the range of 0.01 to 1.0 inch of applied simulated rainfall. The performance of the ETI Noah II gages was similar to the performance of the OTT PLUVIO over much of the range of applied simulated rainfall, including all depths over the range of 0.75 to 8.0 inches. For extremely low amounts of applied simulated rainfall, 0.01 to 0.50 inch, the OTT PLUVIO cumulative absolute differences totals were smaller than those for the other gages tested. Overall, the OTT PLUVIO gages had the lowest cumulative absolute difference totals for up to 2.0 inches of applied simulated rainfall. For 4 to 8 inches of applied simulated rainfall, the ETI Noah II gages had the lowest cumulative absolute differences, followed by the OTT PLUVIO gages. It is also worth noting that the gage currently in use throughout the NADP, the Belfort 5–780, performed better than the Geonor T–200 and Belfort 3200 gages in the Phase I cumulative absolute difference analysis but not as well as the OTT PLUVIO or ETI Noah II gages.

In Phase I testing, the median differences between measured and applied rainfall approximated 0.00 at all depths for the ETI Noah II and OTT PLUVIO gages, as well as for the NovaLynx stick gage, the gage with no moving parts included in the study as a reference gage. For all gages tested, the median percent difference at each depth applied was between –5 and +5 percent. For both units of the ETI Noah II and the OTT PLUVIO gages, the median percent differences were within the narrow range of –0.5 to +0.5 percent.

Following the Phase I indoor bench tests of applied simulated rainfall, all of the gages were relocated outdoors at the Hydrologic Instrumentation Facility for the Phase II testing. Data were collected between January 8 and July 8, 1999. Except for the NovaLynx stick gages, which were read daily, all of the gages were set up to record precipitation continuously. Changes in gage readings due to evaporation losses, offset differences caused by the addition of oil to prevent evaporation, and the dumping of gage contents were taken into account.

Data completeness was different for the various gages during Phase II testing. Both of the OTT PLUVIO gages operated trouble free for the entire 26 weeks of outdoor testing, and weekly records were complete for each of these gages. The Geonor gages experienced only minor losses of data, lasting up to several hours, due to battery failures. The ETI Noah II–1 gage operated for 24 of the 26 weeks, while the ETI Noah II–2 gage operated for 16 of the 26 weeks. Mechanical difficulties with the Belfort 3200–1 and Belfort 3200–2 gages rendered them inoperable for 17 of 26 and 20 of 26 weeks of testing, respectively. Due to operator errors the Belfort 5–780–1 and Belfort 5–780–2 collected data for 25 of 26 and 23 of 26 weeks, respectively. The NovaLynx stick gages were read for the first 25 weeks of the 26-week study.

In 18 of the 26 weeks, the amount of precipitation that fell was between 0.01 and 1.0 inch. The weekly total precipitation was between 1 and 2 inches for 2 of the 26 weeks and between 2 and 3.5 inches in the 2 remaining weeks of the 26-week study. All of the precipitation was in the form of rain. Throughout Phase II testing, the Geonor T–200 weekly totals were frequently slightly higher than the weekly totals for the other technologically advanced rain gages tested due to a few sporadic recordings of trace amounts of precipitation over the course of a week when other gages were recording no precipitation. In 15 out of 20 weeks in which the Geonor T–200 and OTT PLUVIO gages recorded measurable (0.01 inch or more) precipitation, the Geonor T–200 gages recorded a slightly higher weekly total than the OTT PLUVIO gages. The differences in weekly totals between the Geonor T–200 and OTT PLUVIO gages ranged from –0.07 to 0.13 inch, with a median difference of 0.02 inch. The Geonor T–200 gages also recorded slightly higher weekly totals in 15 out of 18 weeks in which the Geonor T–200 and ETI Noah II gages were both operational and recording measurable amounts of precipitation. The differences in weekly totals between the Geonor T–200 and ETI Noah II gages ranged from –0.04 to 0.26 inch, with a median difference of 0.05 inch.

As part of the evaluation of Phase II results, the average weekly precipitation totals obtained from the Belfort 5–780 gages and from each of the gages under consideration as possible replacements for the Belfort 5–780 gage were all compared with the average precipitation weekly totals obtained from two NovaLynx stick gages. The median absolute differences between a particular gage model and the NovaLynx stick reference gage for the 26 weeks of outdoor testing ranged from 0.04 inch for the ETI Noah II and OTT PLUVIO gages to 0.06 inch for the Geonor T–200. The total absolute difference between a particular gage type and the reference gage ranged from 1.23 inches for the Belfort 5–780 to 1.83 inches for the Geonor T–200 gages. Because the Belfort 3200 gages were inoperable for most of the Phase II testing, it is not meaningful to include the results from that gage type in a calculation of median or total absolute differences. The Wilcoxon signed-rank test results indicate there were no statistically significant differences between results from any of the rain gages and results from the NovaLynx stick gages with the exception of the ETI Noah II gages ($p = 0.04$, $\alpha = 0.05$).

The results of the Phase II testing also were analyzed to determine the precision between gages of identical type. Gages of the same type ideally will record the same precipitation amounts. The median relative differences between gages of the same type ranged from 0.00 to 0.01 inch of precipitation for all of the gages except the Belfort 3200, which had operational difficulties. The total absolute difference between gages of the same type over the 26 weeks of the study was 0.12 inch for the NovaLynx stick gages, 0.22 inch for the OTT PLUVIO gages, 0.28 inch for the ETI Noah II gages, 0.51 inch for the Belfort 5–780 gages, indicating good precision between measurements by paired gages of the same type. The total absolute differences for the Geonor T–200 and Belfort 3200 gages were 0.75 and 0.87 inch, respectively. In all cases, the null hypothesis that there was no difference between the paired gages of the same type and model was not rejected.

REFERENCES CITED

Ashmore, S.E., 1934, The splashing of rain: London, Quarterly Journal of the Royal Meteorological Society, v. 60, no. 257, p. 517– 522.

Belfort Instrument Company, 2002a, Universal Precipitation Gauge, Series 5–780/5915; accessed October 24, 2002, at URL http://www.belfortinstrument.com/products/precipitation/

Belfort Instrument Company, 2002b, Glossary: accessed December 9, 2002, at URL http://www.belfortinstrument.com/glossary.html

Belfort Instrument Company, 2002c, Hi-capacity precipitation gauge, model 3200: accessed December 9, 2002, at URL http://www.belfortinstrument.com/products/precipitation/

Bruce, J.P., and Potter, J.G., 1957, The accuracy of precipitation measurements: Toronto, Ontario, Proceedings of the 3d National Meeting of the Royal Meteorological Society, Canadian Branch, June 1957, p. 1–15.

Claybrooke, R.D., Bowersox, V.C., and Lynch, J.A., 2000, Modernizing NTN equipment, a look at two candidates—The NOAH II Precipitation Gage and the Canadian MIC Precipitation Collector: Proceedings, 93d Air and Waste Management Association Meeting, Salt Lake City, Utah, June 19–21, 2000, Paper 389.

Emerson, D.G., and Macek-Rowland, K.M., 1990, Solid-precipitation (snowfall) measurement intercomparison, Bismarck, North Dakota: U.S. Geological Survey Fact Sheet 90–124, 2 p.

ETI Instrument Systems, Inc., 2002, Products: accessed October 24, 2002, at URL http://www.nohowinc.com/eti.htm

Geonor, Inc., 2002, Leakage and precipitation monitoring: accessed December 9, 2002, at URL http://www.geonor.com/leakage.html

Gold, E., 1931, The splashing of rain: London, The Meteorological Magazine, v. 66, no. 787, p. 153–160.

Golubev, V.S., 1985, On the problem of stand conditions for precipitation gage installation, *in* Sevruk, B., ed., Correction of precipitation measurements: World Meteorological Organization/IAHS/ETH Workshop on Precipitation Measurement, St. Moritz, Switzerland, December 2–5, 1993, p. 181–182.

Golubev, V.S., Groisman, P.Y., and Quayle, R.G., 1992, An evaluation of the U.S. 8-inch nonrecording rain gage at the Valdai polygon, Russia: Journal of Atmospheric Oceanic Technology, v. 49, p. 624–629.

Goodison B.E., Ferguson, H.L., and McKay, G.A., 1981, Comparison of point snowfall measurement techniques, *in* Gray, D.M., and Male, M.D., eds., Handbook of Snow: Tarrytown, N.Y., Pergamon Press, p. 200–210.

Groisman, P.Y., and Legates D.R., 1994, The accuracy of United States precipitation data: Bulletin of American Meteorological Society, v. 75, no. 2, p. 215–227.

Handock, D.E., 1960, The measurement of rainfall: Melbourne, Australia Bureau of Meteorology, Seminar on Rain, paper no. 6, p. 1–7.

Hanson, C.L., 1989, Precipitation catch measured by the Wyoming shield and the dual-gauge system: Water Resources Bulletin, v. 25, no. 1, p. 159–164.

Helsel D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: New York, Elsevier Science Publishing Company, Inc., 522 p.

Karl, T.R., Quayle, R.G., and Groisman, P.Y., 1993, Detecting climate variations and change—New challenges for observing and data management system: Journal of Climatology, v. 6, no. 8, p. 1481–1494.

Larkin, H.H., Jr., 1947, A comparison of the Alter and Nipher shields for precipitation gages: Bulletin of American Meteorological Society, v. 28, no. 4, p. 200–201.

Larson, L.W., and Peck, E.L., 1974, Accuracy of precipitation measurements for hydrological modeling: Water Resources Bulletin, v. 10, no. 4, p. 857–863.

Legates, D.R., and DeLiberty, T.L., 1993, Precipitation measurement biases in the United States: Water Resources Bulletin, American Water Resources Association, v. 29, no. 5, p. 855–861.

Metcalfe, J.R., Ishida, S., and Goodison, B.E., 1994, A corrected precipitation archive for the Northwest Territories, *in* Cohen, S.J. (ed.), 1994, Mackenzie Basin Impact Study—Interim Report no. 2, Proceedings of the Sixth Biennial AES/DIAND Meeting on Northern Climate and Mid-Study Workshop of the MacKenzie Basin Impact Study, Yellowknife, Northwest Territories, April 10–14, 1994: Downsview, Ontario, Environment Canada, p. 110–117.

Metcalfe, J.R., Routledge, B., and Devine, K., 1997, Rainfall measurement in Canada—Changing observational methods and archive adjustment procedures: Journal of Climatology, v. 10, no. 1, p. 92–101.

National Atmospheric Deposition Program, 2002, NADP/NTN cooperators: accessed December 8, 2002, at URL http://nadp.sws.uiuc.edu/sponsors.asp

NovaLynx Corporation, 2002, 260-2510 Standard rain and snow gauge: accessed October 24, 2002, at URL http://www.novalynx.com/260-2510.html

Sturges, D.L., 1984, Comparison of precipitation measured in gages protected by a modified Alter shield, Wyoming shield and stand of trees: Proceedings of the 52d Meeting of the Western Snow Conference, p. 57–67, Sun Valley, Idaho, April 1984.

Yang, Daqing, Goodison, B.E., Metcalf, J.R. Louie, P.Y., Leavesley, G.H., Emerson, D.G., Hanson, C.L., Golubev, V.S., Esko, E., Gunther, T., Pangburn, T., Kang, E., and Milkovic, J., 1999, Quantification of precipitation measurement discontinuity induced by wind shields on national gauges: Water Resources Research, v. 35, no. 2, p. 491–508.

Gordon—EVALUATION OF CANDIDATE RAIN GAGES FOR UPGRADING PRECIPITATION MEASUREMENT TOOLS FOR THE NATIONAL ATMOSPHERIC DEPOSITION PROGRAM

USGS/WRIR 02–4302